

# A Theory of Historical Discovery: The Construction of Componential Models

JAN M. ZYTKOW

*Department of Computer Science, Wichita State University, Wichita, Kansas 67208, U.S.A. and  
The University of Warsaw*

HERBERT A. SIMON

*Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, U.S.A.*

(Received August 1, 1985)

**Key words:** scientific discovery, componential models, recovering from inconsistencies, history of chemistry, phlogiston theory

**Abstract.** One of the major goals of 18th century chemistry was to determine the components of substances. In this paper we describe STAHL, a system that models significant portions of 18th century reasoning about compositional models. The system includes a number of heuristics for generating componential models from reactions, as well as error recovery mechanisms for dealing with inconsistent results. STAHL processes chemical reactions incrementally, and is therefore capable of reconstructing extended historic episodes, such as the century-long development of the phlogiston theory. We evaluate STAHL's heuristics in the light of historical data, and conclude that the same reasoning mechanisms account for a variety of historical achievements, including Black's models of mild alkali and Lavoisier's oxygen theory. STAHL explains the generation of competing accounts of the same reactions, since the system's reasoning chain depends on knowledge it has accumulated at earlier stages.

## 1. Introduction

The 18th century was an active one for chemistry, starting with qualitative models such as the phlogiston theory of combustion and ending with the first quantitative models and Dalton's atomic theory. Although the theories of 18th century chemistry were relatively simple, their history is rich in disputes among proponents of different accounts that were designed to explain the same experiments. Several leading theories went through remarkable sequences of transformations in response to being challenged by new experimental data. As a result, this period is an ideal testing ground for theories of scientific discovery, and we will focus on it in this paper. Since we cannot hope to account for the full range of scientific behavior even during this

early stage in the development of chemistry, we will further limit our attention to one facet that was a central concern to chemists of the period – the formulation of componential models.

One of the primary goals of 18th century chemistry was to determine the *components* of various substances. A typical example of such a componential model is ‘marine acid consists of inflammable air and chlorine.’<sup>1</sup> However simple such componential models were, it was very difficult to construct a coherent set of such models for all the substances known by the end of the 18th century. One is struck by the great variety of alternative componential models that were devised during this period and by the long-lasting disputes among their proponents. For example, early chemists produced two quite different sets of models for basic combustion processes – the first set was collectively called the phlogiston theory, while the second set (which was developed several decades later and finally replaced the first) was known as the oxygen theory. Both theories encountered several major obstacles, but for decades their adherents found ways to modify the theories while retaining the basic tenets.

In the following pages we present STAHL, a chemical discovery system that generates componential models like those developed during the 18th century. In the next section we describe the system in terms of its inputs and outputs, as well as in terms of the rules of inference it employs. Since STAHL is intended as a serious model of historical discoveries, we also consider the evidence for the system’s rules and reasoning strategies, presenting examples from the development of the phlogiston theory that they explain. After this, we discuss the control structure used by the program to direct its search through the space of componential models, along with its methods for recovering from errors in reasoning. We then turn to other aspects of 18th century chemical reasoning, showing that STAHL’s explanatory ability is not limited to the phlogiston theory. Finally, we consider some limitations of the current system, and propose some directions for future research. However, before describing our model of early chemical discovery, let us first consider some criteria by which such models should be judged.

### *1.1 Testing models of historical discovery*

In the history of science, alternative accounts of the same data have frequently been proposed, and this suggests an important requirement on computational models of the discovery process – such models should be able to arrive at plausible laws or

<sup>1</sup> In today’s terminology, marine acid is hydrochloric acid (HCl), and inflammable air is hydrogen (H). We have retained the original terminology to avoid the use of cues that the 18th century chemists did not possess, and hence to provide additional insight into their ways of thinking. While we may find it difficult to think in the language of 18th century chemistry, the computer system we shall describe makes no distinction between modern and traditional terminology.

explanations even if these were ultimately rejected in favor of others. The ability to reconstruct different and even competing frameworks, especially ones that persisted over long periods, is an important test of the historical and psychological adequacy of a theory of discovery.

Much of the earlier work on computational models of discovery (including our own) explained only how discoveries *might* have been made and laid no claims to historical accuracy. Although the successful rediscovery of a concept by Lenat's (1977) AM or a law by Langley's (1981) BACON certainly demonstrates one path to that discovery, it does not demonstrate that this path was taken historically. One way to ensure greater historical validity is to require that one's model account for a historical *sequence* of discoveries, rather than isolated events. This will provide a much stronger test for the theory of discovery.

What historic evidence should be used to test whether a theory of discovery successfully models the scientific thinking in a particular epoch? This is not an easy question. Is every instance of scientific reasoning relevant to testing the system? Perhaps not, for a scientist may make an error in a particular application of his method, and we cannot expect to simulate all such mistakes. But conclusions, even mistaken ones, that were accepted over a long period of time and by several leading scientists are likely candidates for confirmatory tests. If one's model of discovery aims at grasping the main currents of reasoning in a given epoch, then reproducing the errors that were typical of that epoch is diagnostic. Here 'errors' means conclusions that were later rejected, but that temporarily enjoyed a widespread acceptance. If we aim at reconstructing the method of reasoning during some period, we should demonstrate that our model accounts for the systematically collected evidence of scientific practice during that period. In this way we ensure that the evidence was not chosen arbitrarily, and that the system's reconstruction is not ad hoc, fitting only selected data points.

An *incremental* discovery system is particularly well-suited for this type of evaluation. Our tests of the model can then closely follow the history of science. We can provide the system with data piecemeal, in the order in which experiments were actually made in history. Then we can see how the inferences the system derives from earlier evidence interact with the later evidence. This technique provides a strong test of the reconstruction, for the 'input' for later discoveries is created by the system itself, not by the programmer. In this way we may protect ourselves against the unconscious (and conscious) introduction of ad hoc assumptions.

A system that models historical developments can serve several purposes. Obviously, it can answer many specific questions about a particular epoch. In the case of 18th century chemistry these include notorious questions about the phlogiston-oxygen conflict. For example, was the phlogiston theory a mistake resulting from inadequate thinking and maintained by blindness or conservatism, or was it a plausible set of inferences given the available data? What (if anything) was superior in Lavoisier's method, and what accounted for the final success of his oxygen theory?

Second, a detailed model of discovery can be a source of analogies for constructing similar systems. For example, if a computational model incorporates specific procedural criteria for choosing among multiple interpretations of data, these criteria can then be applied to other domains. Each success provides additional confirmation for the model of scientific choice. Analogous extrapolation may be attempted with criteria for recognizing inconsistencies, for recovering from errors, and so forth.

Third, simulations of historic episodes provide very concrete evidence about the evolution of scientific method. Discussions about scientific methods are plagued by inconclusive and vague understanding of the scientific method during a given epoch. If methods can be specified at the precise level of procedures that enter into operative simulation programs, it becomes clear what methods are being used at any given time, and how these methods change over time.

### *1.2 Beginnings of the phlogiston theory*

The phlogiston theory originated near the end of the 17th century. It was developed by the German chemist G.E. Stahl (1660–1734), who adopted the ancient view that fire, heat, and light are different manifestations of a common principle that leaves a body during combustion. He called this principle ‘phlogiston’ and initiated an extensive use of the concept in reasoning about chemical reactions. After undergoing several transformations, the phlogiston theory was widely accepted until the 1780’s, when it was rapidly supplanted by the oxygen theory.<sup>2</sup>

In the phlogiston theory, any reaction involving combustion was viewed as a decomposition of the burned body. For instance, Stahl interpreted the burning of charcoal as its decomposition into the matter of fire (another term for phlogiston) and ash.<sup>3</sup> Early phlogistians were not able to isolate phlogiston, but the generation of fire during combustion seemed to be a good observational reason for admitting the production of a substance from the burning body. Later, the existence of phlogiston was supported by a considerable body of evidence as this substance proved useful in explaining many additional reactions.

One of the early successes of the phlogiston theory was its explanation of the smelting of iron from iron ore (calx-of-iron) and charcoal when the latter substances were heated together. According to the phlogistians, this well-known process involved the decomposition of charcoal into phlogiston and ash, followed by the combination of phlogiston with calx-of-iron to form iron. Similar explanations were pro-

<sup>2</sup> The interested reader may wish to consult the following sources on the history of the phlogiston theory: Koertge (1969), Musgrave (1976), Partington (1961, 1962), Zytkow and Lewenstam (1982).

<sup>3</sup> Several decades later, in the second half of the 18th century, fixed air (carbon dioxide) was discovered and recognized as the product of burning charcoal.

posed for the other metals that could be reduced from their calxes when heated in the presence of charcoal.

## 2. The STAHL system

Having reviewed one episode from the initial development of the phlogiston theory, let us now consider a computational model of this development. Our model is implemented as a running computer program named STAHL, after the originator of the phlogiston theory. We will describe the system at a number of levels – in terms of its inputs and outputs, in terms of its intermediate data representation and the rules it employs, and in terms of the control structure for applying those rules.

### 2.1 STAHL's inputs and outputs

The STAHL system accepts an ordered list of chemical reactions as input, and generates as output a list of chemical elements together with the compounds in which they are components. For STAHL, as for human chemists, 'being an element' is a property relative to the chemical reactions available. Thus, from descriptions of chemical reactions it is given,<sup>4</sup> STAHL derives componential models of the substances involved in those reactions. The program may be viewed as a chemist-theoretician who considers empirical data and postulates components for the chemicals involved in the data. In doing this, the program addresses the same task as the human Stahl (1730), for whom

“Universal chemistry is the Art of resolving mixt, compound, or aggregate Bodies into their Principles; and of composing such Bodies from those Principles.”

Stahl's concern was shared by his colleagues and followers. The constitution of substances was a vital problem to every chemist of the 18th century, and most of the disputes and theoretical developments in that century focused on componential models.

Like Langley, Zytkow, Simon, and Bradshaw's GLAUBER (1986), the STAHL program accepts qualitative facts as input, and generates qualitative statements as output. However, STAHL's conclusions can be viewed as explanations of the structure of individual substances and are quite different from the descriptive summaries (generalizations) produced by GLAUBER. The system's initial state consists of a set of reactions, represented in the same schema-like format used by GLAUBER. For

<sup>4</sup> The current version of STAHL does not design its own experiments, though extending the model to include this ability is an obvious direction for future research.

instance, the reaction of burning charcoal in the presence of air would be represented as (reacts inputs {charcoal air} outputs {phlogiston ash air}), in which braces indicate that the order of elements is irrelevant.

STAHL represents its inferences – the components of non-elemental substances involved in the given reactions – in the same formalism as the initial reactions. Thus, the conclusion that charcoal is composed of phlogiston and ash would be stated as the following componential model: (components of {charcoal} are {phlogiston ash}). Intermediate states of STAHL's computation consist of transformed versions of the initial reactions, with the inferences about the components of some substances.

## 2.2 *The heuristics of STAHL*

STAHL incorporates several inference rules for analyzing chemical reactions and building componential models. The most basic of these rules deals with simple synthesis and decomposition reactions, and lets the system unambiguously infer the components of a compound. It can be stated:

INFER-COMPONENTS  
If A and B react to form C,  
or if C decomposes into A and B,  
then infer that C is composed of A and B

This rule can be used to determine the components of charcoal. Given the information that charcoal decomposes to form phlogiston and ash, STAHL would infer that the first substance is composed of the latter two. Note that STAHL does not draw any conclusions about the *amount* of phlogiston and ash contributing to charcoal, but only that they contribute something. Of course, the INFER-COMPONENTS rule is not limited to reactions involving pairs of elements, but can also deal with cases in which three or more substances unite to form a single compound. In the special case of single substances occurring on both sides of a reaction, STAHL infers that they are identical.

If all chemical reactions were of the form shown above, STAHL's task would be simple indeed. However, more complex reactions were common even in the early days of chemistry, and STAHL includes additional rules for dealing with them. The purpose of these rules is to transform descriptions of complex reactions so they can eventually be matched by the INFER-COMPONENTS rule shown above. One such operator is responsible for 'canceling' out substances occurring on both sides of a reaction. The reduction heuristic which proposes this inference can be paraphrased:

**REDUCE**

If A occurs on both sides of a reaction,  
then remove A from the reaction

This heuristic produces a simplified version of a reaction. For instance, if STAHL is given the more complete description of the charcoal burning reaction (reacts inputs { charcoal air } outputs { phlogiston ash air }), the REDUCE rule would apply, giving the simplified reaction (reacts inputs { charcoal } outputs { phlogiston ash }). This revised relation would then be used by the INFER-COMPONENTS rule to infer that charcoal is composed of phlogiston and ash.

A third rule that STAHL incorporates leads initially to more complex statements of reactions, but may make it possible to apply the REDUCE rule. It can be stated:

**SUBSTITUTE**

If A occurs in a reaction,  
and A is composed of B and C,  
then replace A with B and C

The SUBSTITUTE rule draws on information about the components of a substance that have been inferred earlier. For instance, the system may know that charcoal is composed of phlogiston and ash, and that (reacts inputs { calx-of-iron charcoal } outputs { iron ash }). In this case, the SUBSTITUTE rule would let STAHL rewrite the second relation as (reacts inputs { calx-of-iron phlogiston ash } outputs { iron ash }). Given this formulation, the REDUCE rule would lead to (reacts inputs { calx-of-iron phlogiston } outputs { iron }), and the INFER-COMPONENTS rule would conclude that iron is composed of calx-of-iron and phlogiston. As before, the SUBSTITUTE rule is not restricted to substances composed of two elements, but works equally well for more complex structures.

*2.3 An example: the early phlogiston theory*

Let us summarize the above inferences by examining the whole path taken by STAHL in arriving at the componential models of charcoal and iron as postulated by the early phlogiston theory. We present the system with two facts:

(reacts inputs { charcoal air } outputs { phlogiston ash air })  
(reacts inputs { calx-of-iron charcoal air } outputs { iron ash air }).<sup>5</sup>

<sup>5</sup> Calx of iron was the historic name for iron oxide; we have used the original terminology because the modern term is based on the oxygen theory developed by Lavoisier.

One may question the exact representation of these facts, but clearly something very much like this was believed during the period in which the phlogiston theory was developed. We will address the inputs presented to the system in more detail below.

Given the above information, STAHL immediately applies its REDUCE operator to the first fact, giving the revised reaction (reacts inputs {charcoal} outputs {phlogiston ash}). This revision, combined with the INFER-COMPONENTS rule, leads to the inference that charcoal is composed of phlogiston and ash, which was one tenet of the early phlogiston theory. The system then applies its REDUCE operator to the second fact, giving the reduced reaction (reacts inputs {calx-of-iron charcoal} outputs {iron ash}). Having arrived at this conclusion, STAHL applies the SUBSTITUTE rule, generating the expanded relation (reacts inputs {calx-of-iron ash phlogiston} outputs {iron ash}). At this point, the REDUCE rule is used to remove ash from both sides of the equation, giving (reacts inputs {calx-of-iron phlogiston} outputs {iron}). Finally, the INFER-COMPONENTS operator leads STAHL to infer that iron is a compound composed of calx-of-iron and the matter of fire. Table

*Table 1.* STAHL's steps in formulating the phlogiston model

---

Initial state S1:
(reacts inputs {charcoal air} outputs {phlogiston ash air})
(reacts inputs {calx-of-iron charcoal air} outputs {iron ash air})
REDUCE leads to state S2:
(reacts inputs {charcoal} outputs {phlogiston ash})
(reacts inputs {calx-of-iron charcoal air} outputs {iron ash air})
INFER-COMPONENTS leads to state S3:
(components of {charcoal} are {phlogiston ash})
(reacts inputs {calx-of-iron charcoal air} outputs {iron ash air})
REDUCE leads to state S4:
(components of {charcoal} are {phlogiston ash})
(reacts inputs {calx-of-iron charcoal} outputs {iron ash})
SUBSTITUTE leads to state S5:
(components of {charcoal} are {phlogiston ash})
(reacts inputs {calx-of-iron phlogiston ash} outputs {iron ash})
REDUCE leads to state S6:
(components of {charcoal} are {phlogiston ash})
(reacts inputs {calx-of-iron phlogiston} outputs {iron})
INFER-COMPONENTS leads to final state S7:
(components of {charcoal} are {phlogiston ash})
(components of {iron} are {calx-of-iron phlogiston})

---



1 summarizes the states visited by the system in arriving at these conclusions, along with the operators used to generate them.

#### *2.4 Comments on STAHL's inputs*

Before we discuss other models formulated by STAHL, we should address the representations of reactions provided to the system in our examples. We have been careful to use representations for reactions that have some justification in the literature of the period. For instance, by explicitly listing phlogiston as one of the products of combustion, we follow the understanding of combustion by early chemists. Disengagement of fire was such a clear phenomenon that there was little doubt at the time that some form of matter left the combustible and escaped in the form of fire. Lavoisier agreed, too, that a substance disengages during combustion. He called this substance 'caloric' and he believed that it came from oxygen instead of the combustible substance.

After chemists began to study combustion within closed vessels, they realized that air was necessary for combustion to occur. However, they did not at first believe that air changed its chemical identity during this process. Rather, they decided that air played an auxiliary role, similar to that played by water in reactions involving acids, alkalis, and salts. Thus, even starting with empirically more complete descriptions of combustion, such as 'in the presence of air, carbon burns to release phlogiston and to form ash,' they disregarded air in the analysis of this reaction. This inference can be explained by STAHL's reduction heuristic, which would remove the occurrence of air from both sides of the reaction. As a result, the system makes similar 'errors' in reasoning, providing a simple account of the process by which chemists developed phlogiston-based models of combustion. Such confusions are common in the history of chemistry; a similar error, related to the presence of water, led the followers of Lavoisier (around 1810) to believe that sodium was a compound of soda and hydrogen.

The phlogiston chemists noted that some residual ash was left after the combustion of charcoal. Since a similar residuum was the product of any combustion, they inferred that phlogiston could never be found in pure form. (For Stahl, the purest donor of phlogiston was the soot from burning turpentine). Thus, we will include a residual ash in all of our descriptions of early 18th century reactions. Later in the century, as the samples of chemicals became purer, the 'observed' reactions took on a different form, in which most of the residual substances were disregarded.

### 2.5 Incremental processing of the data

One of STAHL's characteristic features is the manner in which its heuristics interact. Note that the SUBSTITUTE rule requires knowledge of a substance's composition, so that some inferences about composition must be made before it can be used. However, we have also seen that complex reactions must be rewritten by the REDUCE and SUBSTITUTE rules before composition inferences can be made. This interdependence leads to a 'bootstrapping' effect, in which inferences made by one of the rules enable further inferences to be drawn; these allow additional inferences, and so forth, until as many conclusions as possible have been reached. This process generally begins with one or more simple reactions to which INFER-COMPONENTS is applied, but after this the particular path taken depends on the data available to the system.

There is no limit to the length of the list of reactions that STAHL can work on, and after the list has been processed, STAHL can take additional data and draw further conclusions from them, repeating this process at will. The more conclusions STAHL has collected, the more powerful its reasoning becomes, as it applies the componential models it has inferred to the analysis of other reactions. This incremental processing makes it possible to model long episodes in the history of chemistry. In our examples, the packages of data input to STAHL are ordered historically to reflect the growing experimental capabilities of chemistry throughout the 18th century. Let us consider some of the other phlogiston-related inferences replicated by STAHL.

### 2.6 Another example: phlogiston in lead and sulfur

We have seen that early chemists believed that all combustible bodies contained phlogiston, since they emitted fire and heat upon burning. In addition, they believed that any metal produced in reactions similar to that of charcoal and calx-of-iron also contained phlogiston. STAHL makes analogous inferences for each of these reactions. For example, suppose we give the system the following situation involving litharge,<sup>6</sup> together with the results from Table 1:

(reacts inputs {charcoal litharge} outputs {lead ash})

Given this additional knowledge, STAHL would apply the SUBSTITUTE rule to replace charcoal with its components, followed by REDUCE to cancel out both occurrences of ash. Finally, the system would use INFER-COMPONENTS, to con-

<sup>6</sup> Litharge is an oxide of lead, with the chemical formula PbO.

clude that lead is composed of litharge and phlogiston.

STAHL also models the phlogistians' belief that sulfur consisted of phlogiston and vitriolic acid. To demonstrate this, G.E. Stahl (Partington, 1961) refers to the following reactions:<sup>7</sup>

(reacts inputs {vitriolic-acid potash} outputs {vitriolated-tartar})  
 (reacts inputs {sulfur potash} outputs {liver-of-sulfur})  
 (reacts inputs {vitriolated-tartar charcoal} outputs {liver-of-sulfur})

Knowing already that charcoal contains phlogiston, STAHL reaches the same conclusion about the composition of sulfur as did the human Stahl: namely, that sulfur consists of vitriolic acid and phlogiston (and ash, unless we see it as one of the products of the last of the above reactions, or disregard it, or use soot as an almost pure source of phlogiston). In this example, the system applies the INFER-COMPONENTS rule three times, and after reaching the second decomposition of liver-of-sulfur, applies SUBSTITUTE to the final reaction, giving:

(reacts inputs {sulfur potash} outputs {vitriolated-tartar charcoal})

At this point the system applies SUBSTITUTE, replacing first vitriolated-tartar and then charcoal with their components. Finally, STAHL employs REDUCE to eliminate potash from both sides of the reaction, obtaining the conclusion that sulfur is composed of phlogiston, vitriolic acid, and ash.

### 2.7 Identification heuristics

Although STAHL's three basic rules – INFER-COMPONENTS, SUBSTITUTE and REDUCE – account for many of the inferences that led to the phlogiston theory, they do not cover all such conclusions. The history of chemistry abounds with cases in which a substance was discovered in two different contexts, was originally thought to be two distinct substances, and was eventually identified as a single substance. STAHL employs two heuristics to model this form of reasoning, in which two substances, originally thought to be different, are assumed to be identical.<sup>8</sup> The first of them may be stated:

<sup>7</sup> In modern terms, vitriolic acid is sulfuric acid, vitriolated tartar is potassium sulfate, and liver of sulfur is a carbonate of sulfur.

<sup>8</sup> The philosophical criterion of Leibniz says that two things are identical if all their properties are the same; for chemical identity it would be enough if two substances had all chemical properties the same. STAHL operates on even weaker criteria of identity, as it does not make use of such properties as color or shape. Does this handicap the system in identifying substances? Not necessarily. Color or shape are not as essential chemical properties as participation in reactions.

**IDENTIFY-COMPONENTS**

If A is composed of B and C,  
 and A is composed of B and D,  
 and neither C contains D nor D contains C,  
 then identify C with D

This heuristic matches when STAHL concludes that a compound can be decomposed in two different ways, but with the decompositions differing by only a single substance. For instance, the identification of phlogiston with inflammable air played an important role in the development of the phlogiston theory. We will see shortly how this identification can be achieved with the rule IDENTIFY-COMPONENTS.

The second heuristic is very similar, except that it applies when two apparently different compounds are found to have the same components. It can be paraphrased:

**IDENTIFY-COMPOUNDS**

If A composed of C and D,  
 and B is composed of C and D,  
 and neither A contains B nor B contains A,  
 then identify A with B

By the application of this rule lime can be chemically identified with chalk and with calcite, as we discuss in the section on Magnesia Alba.

The early chemists acknowledged that changes similar to burning could be produced by the action of acids. Detailed studies of reactions involving solution in acids led to confirming evidence for the phlogiston theory, and eventually, after the discovery of inflammable air (hydrogen) in 1766 by Cavendish, to the identification of inflammable air with phlogiston.

Before we can demonstrate the application of STAHL's identification heuristics in this context, we must consider STAHL's response to the chemical reactions involved in the discovery of inflammable air. Consider the following reactions: (reacts inputs {iron vitriolic-acid water} outputs {vitriol-of-iron inflammable-air water}) and (reacts inputs {calx-of-iron vitriolic-acid water} outputs {vitriol-of-iron water}).<sup>9</sup> Given these facts, STAHL removes the water from both reactions using the REDUCE rule. This sufficiently simplifies the second reaction so that the system can apply the INFER-COMPONENTS rule, concluding that vitriol-of-iron is composed of calx-of-iron and vitriolic-acid. This lets STAHL substitute the components of vitriol-of-iron into the first reaction, giving (reacts inputs {iron vitriolic-acid} outputs {calx-of-iron vitriolic-acid inflammable-air}). After using the REDUCE rule to eliminate vitriolic-acid from both sides of this expression, STAHL infers that iron consists of calx-of-iron and inflammable air.

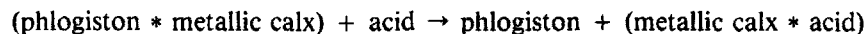
<sup>9</sup> The modern name for vitriol of iron is iron sulfate, while inflammable air is hydrogen.

However, the system knows from the reactions described earlier that iron can also be decomposed into calx-of-iron and phlogiston. Thus, using the first of its identification heuristics (IDENTIFY-COMPONENTS), the system infers that inflammable-air and phlogiston are identical. Both the reasoning and conclusions of STAHL in this example are very similar to those of Cavendish and other phlogiston theorists during the 1760's.

In addition to individual reactions, the 18th century chemists formulated schemes<sup>10</sup> of reactions, such as:

- (A) metal + acid → salt + inflammable air  
 (B) metallic-calx + acid → salt

Metals were considered to be combinations of metallic calxes with phlogiston, and salts were believed to consist of bases and acids. Metallic calxes were classified as bases. From these assumptions, the identification of inflammable air with phlogiston gives the following underlying structure to reaction (A):



where '\*' denotes chemical union, and '+' denotes concatenation of two substances. This schema explained to the phlogistians why the changes produced by acids are similar to calcination and combustion. It also makes clear why no inflammable air effervesces in reaction (B), since there is no phlogiston attached to metallic calx. Although STAHL is unable to generate laws from individual descriptions, it can construct componential models based on general schemes of reactions. Thus, based on (A) and (B) STAHL infers that salt consists of metallic calx and acid, and that metal consists of metallic calx and inflammable air. If, in addition, STAHL knows that metal consists of metallic calx and phlogiston, it concludes that phlogiston is identical with inflammable air.

### 2.8 Evidence for STAHL's heuristics

As we have seen, STAHL's heuristics lead to conclusions very similar to those reached by the chemists of the 18th century. But were these heuristics actually used by those chemists? Let us consider the evidence that they reasoned along these lines.

Our motivation for using the INFER-COMPONENTS rule as the sole basis for structural conclusions relies on the nonambiguity of simple reactions and on em-

<sup>10</sup> Langley et al's (1986) GLAUBER generates just such relational laws from specific reactions, and might be used in conjunction with STAHL to model the generation of these higher level chemical inferences.

pirical access to all the substances involved. Everywhere in the history of chemistry where a synthesis or decomposition reaction was interpreted, its conclusions were generally accepted, unless somebody could show that some 'hidden' reactant had been disregarded. For example, the discovery that calx of mercury decomposed into mercury and oxygen, a reaction that was contrary to the beliefs of phlogistians, forced the phlogistians to alter their theory and to reject the notion that metallic calxes were elements.

However, the conclusions of INFER-COMPONENTS are not as certain when this rule is applied to the description of a reaction to which REDUCE was earlier applied. The REDUCE rule simplifies descriptions of reactions, thus making it possible to draw useful structural conclusions. However, there are situations in which REDUCE produces erroneous conclusions. For example, given the reactions:

- (C) (reacts inputs {copper vitriolic-acid}  
           outputs {sulfurous-acid vitriol-of-copper})  
 (D) (reacts inputs {sulfurous-acid} outputs {vitriolic-acid phlogiston}),

the program infers that:

- (E) copper consists of vitriol-of-copper and phlogiston

Only if provided with knowledge about the composition of vitriol-of-copper (which consists of vitriolic-acid and calx-of-copper) could the system make the correct reduction and infer that:

- (F) copper consists of calx-of-copper and phlogiston.

In other words, the REDUCE rule leads to errors in cases where different amounts of a substance are observed before and after a reaction. From our point of view, such errorful behavior on STAHL's part is desirable, since similar errors were common in the 18th century, and the REDUCE heuristic provides a simple account of their origin.<sup>11</sup> However, these errors were eventually noted and corrected, and STAHL must have a similar resilience if it is to retain historical accuracy. We will consider the system's error recovery methods in a later section.

Using the REDUCE rule, we can model several important cases of reasoning in which similar errors were actually made. Water or air that was present on both sides of reactions was routinely canceled until it became clear that these substances were not only necessary environments of reactions, but had chemical structures that were

<sup>11</sup> One can imagine more conservative versions of the heuristic that require equal amounts of the canceled substance to occur on each side of the reaction. However, since this form of the rule would require quantitative data, it would fail to explain most of the 18th century qualitative inferences.

changed during reaction. Even long after the oxygen-phlogiston dispute had been resolved, at the beginning of the 19th century, we can see examples of this error. After Davy decomposed potash into potassium and oxygen in 1807, Gay-Lussac and Thenard (1808, 1810) argued that potassium is a compound of potash with hydrogen. Their goal was to defend Lavoisier's view of oxygen as the principle of acidity, and given the appropriate data, STAHL repeats their reasoning. From the premises:

(reacts inputs {potassium water} outputs {caustic-potash hydrogen water})  
 (reacts inputs {caustic-potash water} outputs {potassium oxygen})  
 (reacts inputs {potassium ammonia} outputs {hydrogen green-solid})  
 (reacts inputs {green-solid water} outputs {caustic-potash ammonia water})

STAHL reaches the same conclusions as Gay-Lussac and Thenard, that potassium consists of caustic-potash and hydrogen. The second reaction here is a modification of Davy's description (justified by disbelief in the success of Davy's attempts at making caustic potash absolutely dry before applying his source of electricity). In fact, the conclusion of Gay-Lussac and Thenard can be derived from either the first premise, from the second premise (using the assumption that water consists of hydrogen and oxygen), or from the final two reactions.

### 3. The control structure of STAHL

In many cases, STAHL can apply its heuristics in different ways, and this in turn can lead to alternative inferences with different conclusions, and in some cases to inconsistent results and even to infinite recursion. STAHL avoids many of these difficulties by employing a particular control structure, which we consider in detail

*Table 2.* STAHL viewed in terms of search

---

**Initial state:** a list of reactions relating substances

**Goal state:** the components of each compound substance

**Intermediate states:** components of some substances, modified reactions

**Operators/Heuristics:**

**Infer-components:** decides on the components of a substance

**Reduce:** cancels substances occurring on both sides of a reaction

**Substitute:** replaces a substance with its components in a reaction

**Identify-components:** identifies two components as the same

**Identify-compounds:** identifies two compounds as the same

**Search control:** Best-first search with intelligent backtracking

---

below. However, this strategy is not sufficient to avoid all problems, so the system also contains error recovery mechanisms. The emergence of such problems should not be regarded as a deficiency of STAHL, for similar phenomena have been common in the history of chemistry. Often they could be resolved only with the help of new methods of inquiry and new experimental evidence, and STAHL is limited to working with the data it is given. Let us now look at these problems in more detail together with STAHL's response to them.

### *3.1 Testing models for consistency*

STAHL processes one reaction at a time, generating as many inferences as possible from this new information. After the system has applied all of its rules to the reaction, it checks the resulting componential models for internal consistency and for consistency with the componential models that have been accumulated from previous analyses. For instance, STAHL might infer that A consists of B and C *and* that A consists of B, C, and D. Obviously, both componential models cannot be correct. In still other cases, the system may generate an individual conclusion that is internally inconsistent. Two examples of such inferences are (reacts inputs { } outputs { A }) and (reacts inputs { A } outputs { }).

In each of these cases, the principle of conservation of types of substances is violated. The source of these 'inconsistencies' is either an error in the input to STAHL or an inappropriate application of REDUCE. Below we consider STAHL's methods for avoiding such difficulties, and for recovering from them when they occur.

If STAHL fails to construct a componential model based on a given reaction, it drops any intermediate conclusions and remembers the reaction in its original form until new information becomes available. This conservative procedure provides some protection against errors introduced by REDUCE, since more componential models will be available at the later time, lowering the chance of error in both intermediate and final conclusions. Such a practice seems to be historically justified, since the chemical reactions were considered over and over again, and recorded in their originally perceived form rather than in a form partially transformed by their analysis.

If application of REDUCE brings the description of a reaction to the form (reacts inputs { } outputs { }), this means that the input and output of the reaction have canceled each other by reduction. This is a welcome outcome, providing confirmation for the structural knowledge collected by STAHL in the earlier phases of its computation. More precisely, that part of the knowledge that was used in applying the SUBSTITUTE rule to the initial form of the reaction has been confirmed.



### 3.2 Problems arising from multiple inference paths

When several heuristics can be applied to the same reaction, they sometimes produce different results. For example, suppose STAHL knows that (components of {sulfurous-acid} are {vitriolic-acid phlogiston}), and then is given the reaction (reacts inputs {sulfur air} outputs {sulfurous-acid phlogiston air}). Applying the REDUCE rule followed by INFER-COMPONENTS, the system would conclude that (components of {sulfur} are {sulfurous-acid phlogiston}). In contrast, by applying SUBSTITUTE, REDUCE, and INFER-COMPONENTS, it would conclude that (components of {sulfur} are {vitriolic-acid phlogiston}). Both conclusions are correct within the framework of the phlogiston theory. In fact, using the first conclusion together with the premise that sulfurous acid consists of vitriolic acid and phlogiston, one can infer (and thus explain) the second conclusion. However, the reverse is not true; from the second conclusion, one cannot infer, hence cannot explain, the first.

STAHL employs a well-defined control structure to deal with both alternative inference paths and inconsistent results. For every new reaction it encounters, STAHL tries to apply several methods (inference paths), and then to test and compare the results. The first method that STAHL tries to apply is INFER-COMPONENTS. Its result is not a matter of dispute, as the conclusion is based on direct empirical data. If INFER-COMPONENTS cannot be applied, STAHL tries other methods, basically: (a) apply REDUCE first, then INFER-COMPONENTS, (b) apply SUBSTITUTE first, then REDUCE, and finally INFER-COMPONENTS. In fact, the number of alternative conclusions obtained from a reaction may be even greater, since SUBSTITUTE can be applied to different chemicals involved in a given reaction, or can be applied several times in succession. Neither of (a) and (b) is a priori superior over the other. That is, there are cases in which method (a) brings the correct result while the result of application of (b) is incorrect, and cases in which (b) produces the correct conclusion. For this reason, both methods have to be considered whenever possible.

After STAHL has applied all of these methods to the current reaction, it checks the results for consistency with the componential models that have been accumulated from previous analysis. The system then divides these into acceptable results and non-acceptable results, depending on whether they pass the test of consistency. There may also be no result (i.e., no conclusions in the form of componential models) when a method is applied to a reaction. If there are no acceptable results but some that are unacceptable, STAHL enters its automated self-correction procedure, which we will describe shortly. If there is exactly one acceptable result, it will be accepted by STAHL.

If STAHL obtains multiple acceptable results in its analysis of a current reaction, it tries to choose the best of these results. The best result is one (if any), that is acceptable and that explains away all the other acceptable results, based on the following

mechanism. Suppose that R1 and R2 are two different componential models, obtained from a reaction by the use of inference paths P1 and P2 respectively, and D is the set of componential models inferred by the system in the past. In order to determine whether R2 is superior over R1, STAHL substitutes R2 into the current reaction and then tries to apply P1 once again, using D as before. If the reaction is now reduced to empty input and empty output, we may argue that R2 receives an additional confirmation. If the reaction is now transformed into an inconsistent conclusion, we may argue that R2 is to blame. Of course, by symmetry, the second inference path P2 is then examined together with R1. If either R1 or R2 is found to be better, it is accepted and added to the set D of confirmed componential models. If neither R1 nor R2 is found to be better by this procedure, then both inferences are rejected and the reaction that led to them is reconsidered after new data become available.

In view of this procedure we may now reconsider the example of burning sulphur from the beginning of this section. First, STAHL substitutes the second result (components of {sulfur} are {vitriolic-acid phlogiston}) into the sulfur-air reaction. The application of the first inference path (REDUCE followed by INFER-COMPONENTS) ends with the dubious claim that vitriolic-acid is identical with sulphurous-acid. On the other hand, applying the other inference path to the first result (components of {sulfur} are {sulfurous-acid phlogiston}) reduces the initial reaction to empty input and empty output, thereby confirming the premises. Thus, the first result is accepted and the second rejected. In this case, STAHL will later rederive the rejected conclusion by substituting the previously known components of sulphurous-acid (vitriolic-acid and phlogiston) into the componential model for sulfur. Usually, however, the rejected conclusion does not go back to the system.

Let us briefly consider another case of competing inferences, using a reaction involving potassium considered by Gay-Lussac and Thenard, in which (reacts inputs {potassium water} outputs {caustic-potash hydrogen water}). By applying the first inference path, the system infers that potassium consists of caustic-potash and hydrogen. Knowing that water consists of hydrogen and oxygen and applying the second inference path STAHL concludes that potassium is identical with caustic-potash. Now, the system reconsiders both inferences. Using the second result and the first inference path STAHL obtains the inconsistent result (reacts inputs { } outputs {hydrogen}). Using the first result and the second inference path, however, the initial reaction reduces to empty input and empty output. Thus the system rejects the second result and chooses the first one, obtaining Gay-Lussac's model of potassium.

STAHL adds consistent results to its database for use in future reasoning, while it rejects problematic inferences and passes inconsistent results on to its error recovery mechanisms. Although a given reaction may not lead to inferences when first introduced, it is retained in memory and may cause STAHL to draw new conclusions when additional data are added. At each point in its processing, the system's memory contains a list of beliefs about which substances are primitive elements, and about the components of non-elemental substances. When all the reactions have been

considered, some of them may remain unanalyzed; in this case, STAHL applies the same cycle to the remaining data recursively. When no more componential models can be found, the system prints out the structural knowledge it has inferred and halts.

### *3.3 Recovering from inconsistencies*

As we have noted, STAHL recognizes two types of inconsistency – models for the same substances but with different components, and reduced reactions with inputs but no output, or with outputs but no input. Upon realizing that it has generated such an error, the system collects all of the original reactions that contain substances involved in the inconsistency. STAHL then reconsiders the reactions, this time using all the structural knowledge it had already collected, omitting only the inconsistent componential models. Since the knowledge of the system is now larger than at the time when these reactions were considered initially, most of these reactions will quickly reduce to the form of empty input and output. However, this will not occur for all of the data, and usually the resultant componential models will be consistent and the problem resolved.

Let us reconsider the example of vitriol-of-copper discussed earlier, involving reactions (C) and (D). The system inferred that (E) copper consists of vitriol-of-copper and phlogiston, and then, based on other evidence, that (F) copper consists of calx-of-copper and phlogiston. Suppose that still later STAHL concluded that vitriol-of-copper consists of calx-of-copper and vitriolic-acid. This model, combined by inheritance (substitution) with (E), generates a result inconsistent with (F), and this triggers the recovery mechanism. As a result, STAHL reconsiders its knowledge that (C) copper and vitriolic-acid react to form sulfurous-acid and vitriol-of-copper and that (D) sulfurous acid decomposes into vitriolic acid and phlogiston. Based on its new knowledge of the components of vitriol-of-copper, the system concludes that copper consists of calx-of-copper and phlogiston. In other words, (E) is rejected and (F) is retained, because STAHL is able to apply the componential model of vitriol-of-copper, an additional piece of knowledge that was inferred in the meantime.

This process is different from the method of dependency directed backtracking used in some non-monotonic reasoning systems (Stallman and Sussman, 1977; Doyle, 1979). STAHL does not keep track of its inferences, because it models a collective subject that spans over decades rather than a single chemist in time limited to days or weeks. Even a single chemist would have a hard time to remember which inference path he used fifteen years before, and it would often be impossible for him to know the exact way of reasoning of another chemist in Sweden, France, or Germany. Dependency directed backtracking is much more efficient, but science is not organized as neatly as computer databases, and such powerful methods are more difficult to employ.

### 3.4 Dealing with infinite recursion

We have already mentioned the possibility that STAHL may enter an infinite loop or recursion. For instance, given certain reactions involving mercury, calx-of-mercury, and oxygen,<sup>12</sup> STAHL eventually makes two inferences:

(components of {mercury} are {calx-of-mercury phlogiston})  
(components of {calx-of-mercury} are {mercury oxygen})

Taken together, these two inferences imply that mercury is composed of oxygen, phlogiston, and itself. It seems undesirable that an explanatory model should reach such a conclusion.

Infinite recursion does not indicate logical inconsistency: the program applies SUBSTITUTE indefinitely, but at no point will this produce logically contradictory results. Why then do we regard this situation as unsatisfactory? First, an infinite substitution violates our intuitions about the nature of elements, i.e., of the existence of primitive substances from which all others are composed. Second, infinite recursion is technically bothersome. It brings the program to a standstill, and forces the programmer to interrupt processing in an ad hoc manner if he wants the program to continue down more useful paths. Any intervention by the programmer violates the goal of an independently acting computational model.

The recovery method used by STAHL involves four basic components: recognizing the trouble, localizing its source, choosing the recovery procedure, and finally applying the procedure and checking whether the problem has been corrected. As before, we have attempted to implement this process so that it simulates the historical developments in chemistry as closely as possible.

STAHL recognizes cases of infinite recursion by keeping track of the number of levels it has recursed, and exiting after reaching a threshold. Upon exiting, the system collects structural facts that may be responsible for the problem. In the case of the decomposition of red precipitate of mercury, we have seen that these facts are (a) that mercury consists of calx-of-mercury and phlogiston, while (b) calx-of-mercury consists of mercury and oxygen.

Now the system identifies the reactions from which these componential models were drawn and divides the componential models into those obtained by direct inference from simple reactions, and those supported by the use of the SUBSTITUTE-REDUCE technique. The latter facts are more doubtful as being 'less observational' than inferences in the first group. In our example, (b) is recognized as observational, and (a) as a derived fact.

The derived fact is reinterpreted, under a general strategy: 'In case of inconsisten-

<sup>12</sup> Later versions of the phlogiston theory actually included oxygen as an element, but retained phlogiston as their central feature.

cy, make a conceptual distinction'. In our case, the derived component of (a) is the underlying structure of mercury. Then, calx-of-mercury is the only substance recurring in (a) and (b) that is a candidate for reinterpretation. The system replaces it in (a) by the new substance 'calx-of-mercury-proper' (of course, this name is specified by the programmer).<sup>13</sup> As a result (a) is replaced by 'mercury consists of calx-of-mercury-proper and phlogiston', while (b) remains unchanged.

After introducing a conceptual distinction like the one above, STAHL replays the inference process that led to the infinite recursion, to ensure that the alteration has had the desired effect. The system then resumes its main mode and continues processing new reactions.

STAHL's strategy for dealing with infinite recursions reproduces historic lines of reasoning for the case of metallic calxes. Still, it might be an ad hoc solution that we created to deal with this particular problem. We may claim that our solution represents a method characteristic to 18th century chemistry only if the same mechanism deals adequately with other historical cases. Let us consider an episode that occurred three decades later, this time involving the followers of Lavoisier. It will be helpful to recall the reactions of potassium and potash discussed earlier in the paper. Gay-Lussac and Thenard claimed that potassium consists of potash and hydrogen, while Davy demonstrated that by the use of electricity potash can be decomposed into potassium and oxygen. Together these componential models form a troublesome pair analogous to the one involving mercury and its calx that gave the phlogistians difficulty some thirty-five years earlier. The solution of Gay-Lussac and Thenard was also analogous to that developed by the phlogiston theorists. They claimed that in Davy's experiment the potash was not pure, but that it was saturated with water. STAHL applies here the same reasoning as in the case of calx of mercury, inferring that the potash of Davy consists of potash-proper and water.

#### **4. Beyond the phlogiston theory**

Now that we have examined STAHL's behavior in the context of the phlogiston theory, let us see whether the same reasoning can be used to model different episodes in the history of chemistry, such as development of Lavoisier's oxygen theory of combustion. Before considering the oxygen theory, let us consider some work by Black and Berthollet.

<sup>13</sup> In some sense, this process of discrimination is similar to that used by Langley et al's (1986) BACON, which introduces new intrinsic properties when it encounters a situation in which its numerical methods fail to apply. As with BACON, such concepts introduced by STAHL may appear tautological when first introduced, but become respectable to the extent that they prove useful in dealing with other situations besides the one leading to their introduction.

#### 4.1 Black on magnesia alba

In his important work on alkaline substances and fixed air (carbon dioxide), Black (1756) draws several important conclusions about the underlying structures of alkaline substances, and makes an analogy between fixed-air and acids, based on the similar patterns of their reactions with alkalis. We have chosen the following eight reactions to model Black's reasoning:

```
(reacts inputs {lime} outputs {quick-lime fixed-air})
(reacts inputs {quick-lime magnesia-alba14} outputs {lime calcined-magnesia})
(reacts inputs {quick-lime salt-of-tartar} outputs {lime caustic-potash})
(reacts inputs {lime vitriolic-acid} outputs {gypsum fixed-air})
(reacts inputs {magnesia-alba vitriolic-acid} outputs {epsom-salt fixed-air})
(reacts inputs {quick-lime vitriolic-acid} outputs {gypsum})
(reacts inputs {calcined-magnesia vitriolic-acid} outputs {epsom-salt})
(reacts inputs {caustic-potash epsom-salt}
  outputs {calcined-magnesia vitriolated-tartar})
```

Table 3. Inferring the composition of lime and magnesia-alba.

---

```
Initial state S1:
(reacts inputs {lime} outputs {quick-lime fixed-air})
(reacts inputs {quick-lime magnesia-alba} outputs {lime calcined-magnesia}).

INFER-COMPONENTS leads to state S2:
(components of {lime} are {quick-lime fixed-air})
(reacts inputs {quick-lime magnesia-alba} outputs {lime calcined-magnesia}).

SUBSTITUTE leads to state S3:
(components of {lime} are {quick-lime fixed-air})
(reacts inputs {quick-lime magnesia-alba} outputs {quick-lime fixed-air calcined-magnesia})

REDUCE leads to state S4:
(components of {lime} are {quick-lime fixed-air})
(reacts inputs {magnesia-alba} output {fixed-air calcined-magnesia})

INFER-COMPONENTS leads to final state S5:
(components of {lime} are {quick-lime fixed-air})
(components of {magnesia-alba} are {fixed-air calcined-magnesia})
```

---

<sup>14</sup> The modern name for magnesia alba is magnesium carbonate, which has the chemical formula  $\text{MgCO}_3$ .

Let us consider in detail the operation of STAHL's heuristics on the first two of these reactions. The INFER-COMPONENTS rule applies first, leading to the inference that lime ( $\text{CaCO}_3$ ) is composed of quick-lime ( $\text{CaO}$ ) and fixed-air ( $\text{CO}_2$ ). This result enables the SUBSTITUTE heuristic to match, producing a temporarily more complex version of the second reaction, (reacts inputs { quick-lime magnesia-alba } outputs { quick-lime fixed-air calcined-magnesia }). However, since the substance quick-lime occurs in both sides of the modified reaction, the REDUCE rule applies, transforming it into the simpler form (reacts inputs { magnesia-alba } outputs { fixed-air calcined-magnesia }). Finally, this reduced form allows the INFER-COMPONENTS rule to infer that magnesia-alba is composed of fixed-air ( $\text{CO}_2$ ) and calcined-magnesia ( $\text{MgO}$ ). At this point, since no more of its heuristics apply, STAHL concludes that it has formulated as many componential models as the data allow, and halts its operation. The system's behavior on this example is summarized in Table 3.

Upon considering the remaining six reactions and following similar lines of reasoning, STAHL arrives at the following conclusions:

lime	consists of	quick-lime and fixed-air
magnesia-alba	consists of	calcined-magnesia and fixed-air
salt-of-tartar	consists of	caustic-potash and fixed-air
gypsum	consists of	quick-lime and vitriolic-acid
epsom-salt	consists of	calcined-magnesia and vitriolic-acid
vitriolated-tartar	consists of	caustic-potash and vitriolic-acid

Providing STAHL with the additional reactions (reacts inputs { calcite vitriolic-acid } outputs { gypsum fixed-air }) and (reacts inputs { chalk vitriolic-acid } outputs { gypsum fixed-air }) enables the system to draw the conclusions that chalk consists of quicklime and fixed-air, and that calcite consists of quicklime and fixed-air, and then to identify lime with chalk and with calcite by applying its IDENTIFY-COMPOUNDS rule.

In addition to all these componential models, Joseph Black also formulated general laws such as 'a caustic alkali becomes mild after being saturated with fixed air'. In order to generate such laws, STAHL would have to define general classes and qualitative laws that summarized conclusions about individual substances. This form of reasoning is similar to that employed by Langley et al's (1985) GLAUBER, so that one can imagine a version of STAHL combined with GLAUBER to replicate this aspect of Black's reasoning.

#### 4.2 Berthollet on chlorine

In 1774, Scheele added marine acid (hydrochloric acid) to manganese (actually, manganese dioxide), obtaining a brownish liquid. On heating, the liquid became colorless and a gas was expelled, signaling a chemical change that Scheele interpreted as a transfer of phlogiston from the acid to the oxide:

(reacts inputs { manganese marine-acid }  
 outputs { phlogisticated-manganese dephlogisticated-marine-acid })

While Scheele interpreted the expelled gas (which we call 'chlorine') as 'dephlogisticated' marine acid (i.e., marine acid with all phlogiston removed), Lavoisier and his followers viewed the substance as a compound of marine acid and oxygen. Berthollet conducted several experiments, which convinced him that Lavoisier's view was correct. His results can be stated as:

(reacts inputs { chlorine water } outputs { oxymuriatic-acid water })  
 (reacts inputs { oxymuriatic-acid water } outputs { muriatic-acid oxygen water })  
 (reacts inputs { black-manganese } outputs { calcined-manganese oxygen })  
 (reacts inputs { black-manganese muriatic-acid water }  
 outputs { salt-of-manganese chlorine water })  
 (reacts inputs { calcined-manganese muriatic-acid water }  
 outputs { salt-of-manganese water })

When presented with these reactions, STAHL reaches a conclusion similar to that of Berthollet: chlorine consists of muriatic-acid and oxygen. This inference can be drawn from the first two reactions. It can be additionally confirmed by the last three reactions since the fourth reaction will finally reduce to (reacts inputs { } outputs { }) when STAHL applies to it Berthollet's model of chlorine.

#### 4.3 Lavoisier's caloric theory and the discovery of oxygen

In 1773 Priestley experimented with the red calx of mercury (so-called red precipitate) and obtained the element oxygen, which phlogistians later termed dephlogisticated air. This important reaction can be stated:

(reacts inputs { red-calx-of-mercury } outputs { mercury oxygen })

The conclusions of this experiment challenged the claim that metallic calxes are elements. Priestley's result also challenged the phlogiston theorists to identify the source of the phlogiston that produced mercury, the second product of the reaction.



In response to these difficulties, and in response to Cavendish's (1784, 1785) findings on the composition of water, the theory of phlogiston was modified and improved over the years from 1775 to 1785. We have considered the impact of Priestley's experiment in the section about infinite recursion. Interestingly, the same experiment that caused the phlogistians to change their theory leads STAHL into a troublesome infinite loop. STAHL recovers from this crisis in a way very similar to that taken by the phlogiston theorists. STAHL is then able to follow the further development of the phlogiston theory, including the final version proposed by Cavendish (1784, 1785).

Now that we have discussed STAHL's response to these new findings, let us outline the development of Lavoisier's oxygen theory, which he developed during the period in which the phlogiston theory was being adapted to the new discoveries. STAHL can easily reconstruct what is usually accepted as Lavoisier's reasoning on oxides and oxidation. Let us assume the system is given the following reactions:

(reacts inputs {calx-of-lead} outputs {lead oxygen})  
 (reacts inputs {calx-of-lead charcoal} outputs {lead fixed-air})

Given these data, the system infers that calx-of-lead consists of lead and oxygen and that fixed-air consists of charcoal and oxygen. STAHL's ability to reconstruct these conclusions about combustion calls into question Lavoisier's claim that exact, quantitative measurements are both necessary and sufficient to prove the correctness of his theory of combustion and to reject the views of phlogistians. In fact, weighing was not performed as carefully as was claimed by proponents of the quantitative method, either during the culmination of the phlogiston-oxygen dispute in the 1780's or even twenty years later.

The conclusions just cited are not quite those reached by Lavoisier. However, we cannot blame STAHL, since the patterns of the reactions that we gave to STAHL are also not Lavoisier's. For Lavoisier, combustion was a decomposition of oxygen-gas, and his pattern of combustion and reduction was:

combustible + oxygen-gas  $\rightarrow$  oxide + caloric,

where oxygen-gas consists of oxygen-principle and caloric.<sup>15</sup> Calcination of metals was conceived in the same way:

<sup>15</sup> From a phenomenological standpoint (as Lavoisier himself noted) caloric, like phlogiston, evidenced itself as 'matter of fire.' But where the phlogistians found the source of the 'matter of fire' in the combustible or the metal, Lavoisier found it in the oxygen gas. This change of interpretation could not have taken place before oxygen was determined to be an actual participant in (and not a mere condition for) the combustion or calcination process. Nor did the introduction of oxygen into the descriptions of the reactions immediately require a reinterpretation of the source of the 'matter of fire.'

metal + oxygen-gas → metallic-oxide + caloric

where metallic-oxide is a compound of oxygen-principle and metal. The creation of acids was viewed in a similar way. For Lavoisier caloric was an important chemical agent and his descriptions of combustion and calcination reactions all involve caloric:

(reacts inputs {calx-of-lead caloric} outputs {lead oxygen-gas})  
 (reacts inputs {calx-of-lead charcoal caloric} outputs {lead fixed-air})  
 (reacts inputs {charcoal oxygen-gas} outputs {fixed-air caloric})  
 (reacts inputs {water charcoal caloric} outputs {hydrogen-gas fixed-air})  
 (reacts inputs {water iron caloric} outputs {hydrogen-gas oxide-of-iron}).

These data are not sufficient for STAHL to draw any conclusions, and they were indeed not sufficient for Lavoisier to argue that the caloric in the third reaction comes from oxygen-gas rather than from charcoal. Lavoisier's belief that oxygen-gas contains caloric was based on his earlier idea of caloric as the principle of the gaseous form of matter. He noticed that in many reactions in which the input is heated, some kind of air is disengaged, while in reactions in which an air is absorbed, fire or heat is disengaged. In short: Fire in, gas out; fire out, gas in; and in conclusion: Any gas contains caloric.

STAHL has no capability for considering properties like 'gaseous-form', and is not able to generalize about classes of substances. However, if we supply STAHL with statements like: 'Oxygen-gas consists-of oxygen-principle and caloric', then many of STAHL's findings on the composition of substances are the same as Lavoisier's. For example, from this statement and the first reaction above, STAHL concludes that calx-of-lead is composed of lead and oxygen-principle. Lavoisier's reasoning is difficult to model with STAHL's, because the French chemist made extensive use of reasoning based on generalizations (all acids contain oxygen, all gases contain caloric), and on several occasions accepted the conclusions of a generalization rather than facts that contradicted this generalization.

#### 4.4 *The final years of phlogiston*

Between the years 1785 and 1795, most of the supporters of the phlogiston theory abandoned the theory, and accepted the theory of oxygen instead, although they did so to varying extents. Historians and philosophers of science have proposed many explanations of this development (e.g., Musgrave, 1976; Krajewski, 1977; Zytkow and Lewenstam, 1982), but each of these accounts is beyond the scope of the current version of STAHL.

However, STAHL can provide justification of a theory that was proposed during

this transition period and that was meant to satisfy both sides in the conflict, based on the idea that both the phlogiston and oxygen theories embodied part of the truth. Suppose we provide STAHL with the following information: (a) Lavoisier's general pattern for calcination of metals (see the previous section); (b) his belief that metallic-oxide consists of metal and oxygen-principle; (c) his belief that oxygen-gas consists of oxygen-principle and caloric, and (d) the componential model of metals of the phlogiston theory in its version from the 1780's, that is, the claim that metal consists of its proper calx and phlogiston. Upon substituting (b), (c), and (d) into (a), STAHL would produce the following general structure:

$$\begin{aligned} &(\text{metallic-calx-proper} * \text{phlogiston}) + (\text{oxygen-principle} * \text{caloric}) \rightarrow \\ &(\text{metallic-calx-proper} * \text{phlogiston} * \text{oxygen-principle}) + \text{caloric} \end{aligned}$$

which can be reduced to empty input and output, confirming the admissibility of its premises. This is the schema proposed by Gadolin in the 1790s, which fits the empirical data as well as either the oxygen or the phlogiston theory. Although this theory never gained wide acceptance, it shows that there was no inherent reason why the notions of phlogiston and oxygen could not be combined in a single framework.

### 5. Limitations of STAHL

Although STAHL incorporates an important part of 18th century methods for theorizing about the composition of substances, there are several aspects of this period that the system cannot handle. We have already mentioned a number of these. Some of the limitations are easy to remedy, while overcoming others would require substantial additions to the model.

We have seen that the early chemists formulated general laws, such as Lavoisier's statement that combustion involved the combination of combustible substances with oxygen to generate an oxide and caloric. STAHL cannot make such generalizations, but general laws are the main concern of Langley et al's (1986) GLAUBER system, which uses exactly the same format for its data as STAHL. Thus, one can imagine a combined system in which the STAHL component passed its results to the GLAUBER component, from which general laws like Lavoisier's would be formulated. In turn, laws produced by GLAUBER could be passed back to STAHL, which treats general terms like metal, acid, and alkali in exactly the same manner as specific substances like iron, chlorine, or potash. For example, given the schema:

$$(\text{reacts inputs } \{ \text{caustic-alkali fixed-air} \} \text{ outputs } \{ \text{mild-alkali} \})$$

which might be generated by the GLAUBER system, STAHL would conclude that mild-alkali consists of caustic-alkali and fixed-air.

The current version of STAHL does not deal with affinities, which were a very active (and very inconclusive) source of argument during the 18th century. We believe that the ability to make inferences about affinity could be added to STAHL with very little effort. After collecting facts on the internal structure of chemicals, STAHL is in an ideal position to reconsider the reactions already processed, looking for exchange reactions of the form  $(A * B) + C \rightarrow A + (B * C)$ , which are the major source of inferences about affinity. The program need only to be supplemented by one additional heuristic and possibly some adjustments to its control mechanism.

At similarly low cost, STAHL could be supplemented with two other facilities. First, the system could be extended to explain reactions themselves by applying the SUBSTITUTE rule to the description of the reaction, giving forms like that for the exchange relation above. Second, STAHL could be modified to predict the outputs of a reaction from its inputs, using componential models and its knowledge of affinities.

In its current form, STAHL cannot consider properties of substances such as shape, taste or color. There are several episodes from the history of chemistry in which such properties were used in arguing about components of chemicals. For Lavoisier acid taste was a very strong argument for the presence of oxygen in a substance, while aeriform state indicated the presence of caloric. STAHL is not able to reach these conclusions, but can accept them in the form of explicitly added componential models: (components-of {acid} are {acid-radical oxygen-principle}), and (components-of {gas} are {gas-radical caloric}). It can use these models as premises for further inferences. The example of caloric as the principle of elastic fluids was considered earlier.

Another limitation is that the present incarnation of STAHL must rely entirely on data that are provided by an experimenter. In the course of using its error recovery mechanisms, some of these data may be questioned, but the task of their verification and improvement is then passed on to the experimenter again. An experimenter must be able to recognize (identify) substances, and be able to produce a given reaction. He is also responsible for selecting the right amounts of substances so that neither input substance remains in the output. This is essential, since STAHL would fail miserably in dealing with reactions that contained redundant amounts of a substance.

This again raises the issue of quantitative measurement, which was applied by both sides in the phlogiston conflict. However indirectly, the early chemists (and STAHL) used the results of precise weight measurements in order to make their reactions as simple as possible. For example, (reacts inputs {sulfur iron} outputs {sulfur sulfuretted-iron}) can be simplified to the form (reacts inputs {sulfur iron} outputs {sulfuretted-iron}), provided that proper quantities of both sulfur and iron enter the reaction. This process of refining chemical reactions was an important part of the experimental art. During the 1770's and 1780's, supporters of the phlogiston theory used balances no less accurate than Lavoisier's and estimated empirical error in a

realistic way (Cavendish, 1784; Cavendish, 1785; see Nicholson's preface to Kirwan, 1789).

Finally, we should note that STAHL is limited to considering one line of theoretical development at a time, and is not able to compare two different views. The current version routinely considers different inference paths for each reaction, and chooses the best componential model, but does not keep different solutions for further consideration. Therefore it cannot choose between theories, and in particular it cannot model the resolution of the oxygen-phlogiston conflict in favor of either side. Future versions of the system should be able to retain at least a few competing sets of models and test them on their ability to explain new reactions.

## 6. Conclusion

In our earlier work on scientific discovery (Langley et al., 1986) we have claimed only that our systems embody mechanisms that are sufficient for making certain kinds of discoveries. While we have used historical examples to test their capabilities, we have not claimed that they follow closely the original path of historical discovery. In the case of STAHL, however, we believe the claims can go a little farther. The program traces in detail several historical paths of discovery. It provides an explication of alternative conceptualizations of a research domain, and hence of alternative modes of reasoning that characterized competing schools of thought.

STAHL incorporates specific knowledge about the constitution of objects and the conservation of basic substances. It constructs explanations in the form of descriptions of underlying structures of substances and reactions. At the beginning, it is data-driven, but after it has made conjectures about the hidden structures, it is also driven by these conjectures – that is, by theory. It employs general-purpose heuristics, and can use them to choose between multiple conclusions and to deal with some of the inconsistencies in its results.

Applying STAHL to the study of 18th century chemical theorizing, we find considerable consistency in the modes of reasoning used by different chemists, even when they reached different conclusions. In the case of the conflict between the phlogiston and oxygen theories of combustion, it appears that the proponents of the two theories used much the same method of reasoning, differing mainly in their assumptions. The phlogisticians believed 'matter of fire' to be an essential constituent of metals that was driven off during calcination. The oxygen theorists assumed 'matter of fire' (now rechristened 'caloric') to be an essential constituent of 'pure air,' that was driven off when the air combined with combustible material. Because of the difference in assumptions, the same rules of reasoning applied to the same reaction could bring different conclusions.

However, the reasoning embodied in STAHL's heuristics is not peculiar to the theory of combustion. By examining other events of 18th century chemistry, such as

Black's analysis of magnesia alba, we have shown that the same principles of inference were used by chemists quite widely in their search for consistent accounts of the chemical substances and their reactions.

### Acknowledgements

This research was supported in part by Contract N00014-84-K-0345 from the Division of Information Science, Office of Naval Research. We would like to thank Pat Langley and Gary Bradshaw, who contributed significantly to the research described in this paper. Donald Beaver provided many helpful suggestions to an earlier version of this paper.

A modified version of this paper will appear as a chapter in Langley, P., Simon, H.A., Zytchow, J.M., and Bradshaw, G.L., *Scientific Discovery: A Computational Account of the Creative Processes*.

### References

- Berthollet, M. (1897). Memoir on Dephlogisticated Marine Acid. In *The early history of chlorine, alembic club reprints, No. 13*. London.
- Black, J. (1756). Experiments upon Magnesia alba, Quicklime, and some other Alkaline Substances. In *Essays and observations, physical and literary*. . . Edinburgh.
- Cavendish, H. (1766). Three papers, containing experiments of factitious air. *Philosophical Transactions*, 56, 141–184.
- Cavendish, H. (1784). Experiments on Air. *Philosophical Transactions*, 74, 119–153.
- Cavendish, H. (1785). Experiments on Air. *Philosophical Transactions*, 75, 372–384.
- Doyle, J. (1979). A truth maintenance system. *Artificial intelligence*, 12, 231–272.
- Gay-Lussac, L.P., & Thenard, L.J. (1808). Sur les metaux de la potasse et de la soude. *Annales de chimie*, 66, 205–217.
- Gay-Lussac, L.P., & Thenard, L.J. (1810). Observations. *Annales de chimie*, 75, 290–316.
- Kirwan R. (1789). *An essay on phlogiston and the constitution of acids*. London: J. Johnson.
- Koertge, N. (1969). *A study of relations between scientific theories: A test of the general correspondence principle*. PhD thesis, University of London.
- Krajewski, W. (1977). *Correspondence principle and growth of science*. Dordrecht, Holland: Reidel.
- Langley, P. (1981). Data-driven discovery of physical laws. *Cognitive Science*, 5, 31–54.
- Langley, P., Zytchow, J.M., Simon, H.A., & Bradshaw, G.L. (1986). The search for regularity: Four aspects of scientific discovery. In R.S. Michalski, J.G. Carbonell & T.M. Mitchell (Eds.), *Machine learning, Vol. 2*. Los Altos, CA: Morgan Kaufmann Publishers.
- Lavoisier, A. (1789). *Traite elementaire de chimie*. Paris: Chez Cuchet.
- Lenat, D.B. (1977). *Automated theory formation in mathematics. Proceedings of the Fifth International Joint Conference on Artificial Intelligence* (pp. 833–842). Cambridge, Mass.: Morgan Kaufmann.
- Musgrave, A. (1976). Why did oxygen supplant phlogiston? Research programmes in the chemical revolution. In *Method and appraisal in the physical sciences*. Cambridge, England: Cambridge University Press.
- Partington, J.R. (1961). *A history of chemistry*. London: Macmillan.

- Partington, J.R. (1962). *A history of chemistry*. London: Macmillan.
- Priestley, J. *Experiments and observations on different kinds of air*. 3 Vols. London: J. Johnson 1774–5–7.
- Scheele, C.W. (1786). On Manganese, Manganese or Magnesia Vitriariorum. In *The chemical essays of Charles-William Scheele*. London.
- Stahl, G.E. (1730). *Philosophical principles of universal chemistry*. London: Osborn & Longman.
- Stallman, R.M., & Sussman, G.J. (1977). Forward reasoning and dependency-directed backtracking in a system for computer-aided circuit analysis. *Artificial Intelligence*, 9, 135–196.
- Zytkow, J.M., & Levenstam, A. (1982). Czy tlenowa teoria Lavoisiera była lepsza od teorii flogistonowej? przyczynek do analizy rewolucji naukowej. Was the oxygen theory of Lavoisier better than the phlogiston theory? A contribution to the analysis of the scientific revolution. *Studia Filozoficzne*, (9–10), 39–65.