Machine Learning 1: 423–451, 1986

Chemical Discovery as Belief Revision

DONALD ROSE (DROSE @ CIP.UCI.EDU)
PAT LANGLEY (LANGLEY @ CIP.UCI.EDU)
Program in Computation and Learning, Department of Information & Computer Science, University of California, Irvine, CA 92717, U.S.A.

(Received July 15, 1986)
(Revised August 19, 1986)

Key words: machine discovery, belief revision, componential models, chemical reasoning, phlogiston theory

Abstract. In this paper we describe STAHLp, a system that constructs componential models of chemical substances. STAHLp is a descendant of Zytkow and Simon's (1986) STAHL system, and both use chemical reactions and known componential models in order to construct new chemical models. However, STAHLp employs a more unified and effective strategy for recovering from erroneous inferences, based partly on de Kleer's (1984) assumption-based method of belief revision. This involves recording the underlying source beliefs or premises which lead to each inferred reaction or model. Where Zytkow and Simon's system required multiple methods for detecting errors and recovering from them, STAHLp uses a more powerful representation and additional rules which allow a unified method for error detection and recovery. When given the same initial data, the new system constructs the same historically correct models as STAHL, but it has other capabilities as well. In particular, STAHLp can modify data it has been given if this is necessary to achieve consistent models, and then proceed to construct new models based on the revised data.

1. Introduction

Scientific discovery and belief revision are two areas of artificial intelligence which have undergone considerable investigation, yet thus far work in these areas has rarely overlapped. Zytkow and Simon's (1986) STAHL system — which constructs componential models of chemical substances — was a first step toward combining methods from both paradigms. STAHL determines which substances are actually compounds and which substances make up these compounds, employing a simple form of belief revision to resolve conflicts between models and to recover from erroneous inferences. However, we will see that there are some problems with the particular methods the system uses to this end.

In an attempt to improve on Zytkow and Simon's results, we have designed and implemented a successor to their system called STAHLp. Like the original STAHL, the new program accepts a set of chemical reactions as input data and infers componential models of the substances involved in those reactions as its output.[1] Both systems are implemented as forward-chaining production systems, in which production rules match against the beliefs currently residing in working memory. If a rule's conditions match against those beliefs, the rule 'fires' and either asserts new beliefs into working memory or removes existing beliefs. The new state of memory in turn leads to more rule firings and to a revised set of beliefs, leading to componential models of the observed substances.

We will see that there are several differences between the two systems, but the main difference lies in the belief revision capability of STAHLp. One can view a collection of componential models as a theory, and one of our goals was to explore methods for incrementally revising a theory as new data (reactions) are observed. Another motivation was to overcome certain limitations in the original STAHL system. Although Zytkow and Simon's program was able to resolve conflicts among a few beliefs, we wanted a more general mechanism for revising an entire set of beliefs. In short, we hoped to model the processes by which entire theories are revised in response to new information.

Let us consider how STAHLp accomplishes this goal by summarizing the system's basic inference cycle. First, (1) new componential models are generated in a forward-chaining manner until (2) some erroneous inference is noted. The default inference process is then suspended, and the belief revision process in invoked. During this stage, (3) hypotheses are generated that propose modifications of the original input data (premises) to avoid the erroneous inference. Next, (4) the 'best' of these hypotheses is selected and the proposed modifications are carried out. Finally, the system returns to step (1), generating new componential models based on the revised premises, and the cycle continues until a consistent set of models is found. If no errors are noted, only the first step of this cycle is necessary. In fact, this step is itself a cycle in which initial premises lead to intermediate beliefs, which in turn lead to componential models, which are then used to infer more intermediate beliefs, and so on.

In the pages that follow, we describe each of these steps in detail. However, before describing the system itself, we should briefly recount the historical phenomena that STAHL and STAHLp were designed to model. The task domain for both systems is 18th century chemistry, during which qualitative studies of reactions had led to the *phlogiston* theory of combustion. The basic assumption was that burning substances lost something during the process of combustion, and that this substance was phlogiston. The notion of phlogiston also seemed to explain the related problem of *calcination*, during which a metal gradually rusts over time. The 18th century chemists believed that calcination, like combustion,

[1] We will refer to both reactions and models as *beliefs*, but we will reserve the term *premise* for input reactions.

involved the loss of phlogiston. Thus, two problems which had long frustrated chemists suddenly had rational explanations and even seemed to be related phenomena. Stillman (1960) has described the phlogiston theory as the first comprehensive chemical framework. Although eventually proven incorrect, this theory at least gave chemists a foundation on which to build later chemical theories, including Lavoisier's theory of oxygen.

Since STAHLp is similar in many respects to Zytkow and Simon's earlier STAHL program, we begin by describing the basic aspects of these systems, focusing on the representation and rules common to both. In section 3 we shift our attention to the differences between the two systems, discussing some limitations of STAHL and showing how STAHLp's enhanced representation and additional rules let it *prevent* certain classes of errors. In section 4 we will see that this new representation also allows STAHLp to *recover* from other errors through a process of belief revision. Having described the system itself, we then proceed to clarify its operation in section 5 with two examples from the history of chemistry. Finally, we consider the generality of STAHLp's methods, summarizing the historical reasoning it has successfully modeled.

2. Overview of STAHL and STAHLp

Like Zytkow and Simon's STAHL system, STAHLp is a forward-chaining production system which constructs models of substances from chemical reactions provided as input.[2] STAHLp's basic inference cycle starts when a set of reactions are added to working memory. Various production rules match against these beliefs, and upon firing they add new inferences to memory. The goal of any inference chain is to construct a componential model; one can view such a chain as an inverzed pyramid, with premises (reactions) at the top and the new model at the bottom. This new model can then be used in other inference chains to help infer additional models.

2.1 Basic representation

Both STAHL and STAHLp deal with two basic types of beliefs — reactions and componential models. In their basic forms, a reaction is represented as a list of *inputs* and *outputs*, while a model is represented as a list containing a *substance* and its *components*. For example, chemists of the 18th century observed that calx-of-iron[3] reacted with charcoal to form iron and ash; STAHLp would represent

[2] STAHLp's name comes from two sources. Both the original STAHL and the current system are named after the German chemist G.E. Stahl (1660–1734), who originally proposed the phlogiston concept. The 'p' derives from STAHLp's implementation in the PRISM production system language (Langley, Ohlsson, Thibadeau & Walter, 1984), while Zytkow and Simon's program was implemented in LISP.

[3] Today this substance is viewed as an oxide of iron, but like Zytkow and Simon we will use 18th century terminology for discussing and representing substances. This choice of terms does not impact the operation of the system.

this reaction internally as

(reacts inputs {calx-of-iron charcoal} outputs {iron ash})

but in this paper we will use the following more readable notation

calx-of-iron charcoal → iron ash (1)

to represent reactions. Componential models are described in a very similar manner; the program would represent the belief that charcoal is composed of phlogiston and ash as

(components of {charcoal} are {phlogiston ash})

but we will use the more compact description

charcoal = phlogiston ash. (2)

These two types of beliefs have the same conceptual structure; one could say that the 'components' of calx-of-iron and charcoal are iron and ash, but we save this label for when we have inferred the components of just *one* substance, because that is the ultimate goal. Our shorthand notation also suggests an algebraic metaphor; one can view the above reaction and model as 'equations,' and this suggests operations for 'solving' these equations. Thus, one can *substitute* the components of charcoal from equation 2 into equation 1 to get the new reaction:

calx-of-iron phlogiston ash → iron ash. (3)

We can now safely 'subtract' or *reduce* the substance ash from both sides, leaving us with

calx-of-iron phlogiston → iron. (4)

From this reaction we can now *infer* the *components* of iron, thus expanding the number of known models to two in this simple example:

iron = calx-of-iron phlogiston. (5)

Now let us examine in more detail the rules responsible for such reasoning.

2.2 Basic production rules

The three steps just seen correspond to the three main production rules in both

STAHL and STAHLp, which we will call SUBSTITUTE, REDUCE, and INFER-COMPONENTS. We can paraphrase the first two rules as follows:

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

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

where S denotes a set of one or more substances. In general, an application of the SUBSTITUTE rule is followed by application of the REDUCE rule (since substitution can lead to a substance being present on both sides of a reaction). As soon as we generate a reaction with just one substance on either side, we can infer a componential model for that substance. STAHL and STAHLp effectively use the same rule for making these inferences, which we can paraphrase as:

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

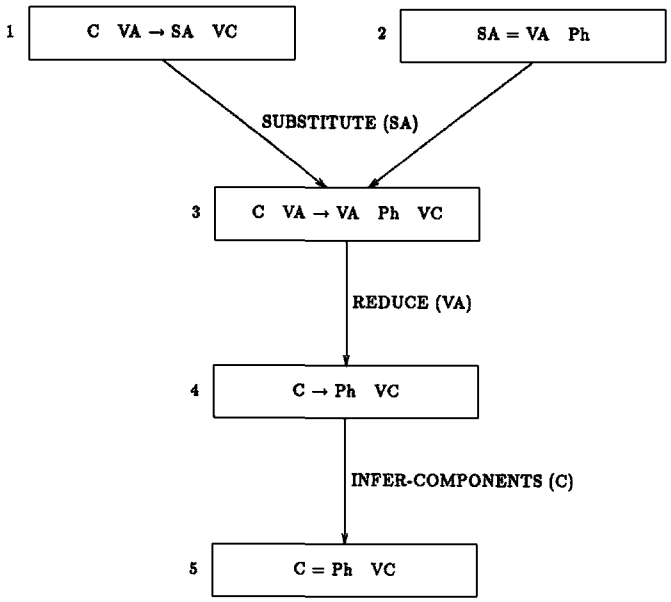

Figure 1. Inferring a model for copper.

At this point, if there exist other reactions in working memory which contain the substance B in their inputs or outputs, the SUBSTITUTE rule can apply again, possibly leading to more firings of REDUCE, to more models being proposed by INFER-COMPONENTS, and so forth. This process continues until no further inferences can be made. STAHL and STAHLp differ slightly in their methods for selecting between competing rules, but we will not focus on the details of the control structure here.

3. Preventing erroneous inferences

While the cycle described above (substitution, reduction, inferring components, further substitution, etc.) works in many cases, occasionally errors occur. For instance, Zytkow and Simon (1986) refer to situations in which 'the REDUCE rule leads to errors' because 'different amounts of a substance are observed before and after a reaction.' In this section we consider the source of such errors and explain how STAHLp's augmented rules and representation let it avoid such problems.

3.1 Problems with STAHL's REDUCE rule

Let us begin with an example from Zytkow and Simon in which STAHL begins with the reaction C VA → SA VC and the componential model SA = VA Ph.[4] Using its basic inference rules, the system generates the revised reaction C VA → VA Ph VC (by substitution), then produces C → Ph VC (by reduction), and finally infers the new model C = Ph VC. Figure 1 presents this reasoning chain in graphic form. However, this conclusion is incorrect even in the phlogiston paradigm; the correct model of copper in this context is C = Ph CC. But one cannot hope to infer the correct version without a missing piece of knowledge — that vitriol-of-copper is actually composed of vitriolic-acid and calx-of-copper (VC = VA CC).

The problem becomes apparent when we consider how STAHL would have utilized the componential model for vitriol-of-copper had it received this information after copper's model was inferred. In this case, the components of vitriol-of-copper would be substituted into the model C = Ph VC (using a related version of SUBSTITUTE for models), giving the new model C = Ph VC CC. Figure 2 shows this alternative chain of inferences.

The reason STAHL cannot infer the correct model (C = Ph CC) is that the system has no mechanism for 'remembering' that VA has already been reduced earlier in the inference chain. When the components of VC are substituted into

[4] We will often abbreviate the names for substances. In this example, C stands for copper, VA for vitriolic-acid, SA for sulfurous-acid, VC for vitriol-of-copper, Ph for phlogiston, and CC for calx-of-copper.

the original model of copper, and VA again appears on the right-hand side, STAHL does not realize it should remove this new occurrence of VA from the model.

Of course, the reader may wonder why VA needs to be removed when it appears again after the substitution. The reason involves a tacit assumption made by Zytkow and Simon (1986) — that all occurrences of a substance on one side of a reaction cancel all occurrences on the other side (i.e., that all occurrences on a side may be condensed into a single occurrence). This is a plausible assumption, since 18th century chemists had not yet taken quantitative measures (such as the conservation of mass) into account in their reasoning. The early chemists were more concerned with whether a substance appeared in the inputs or outputs of a reaction than with how much of that substance was present (which they often found impossible to measure).

Thus, their reasoning (and the reasoning of STAHL) would transform a reaction such as C VA → VA Ph VA CC into C → Ph CC, because the VA on the left would cancel both occurrences of VA on the right. However, the problem is that such reactions rarely appear in memory during STAHL's inferencing; the

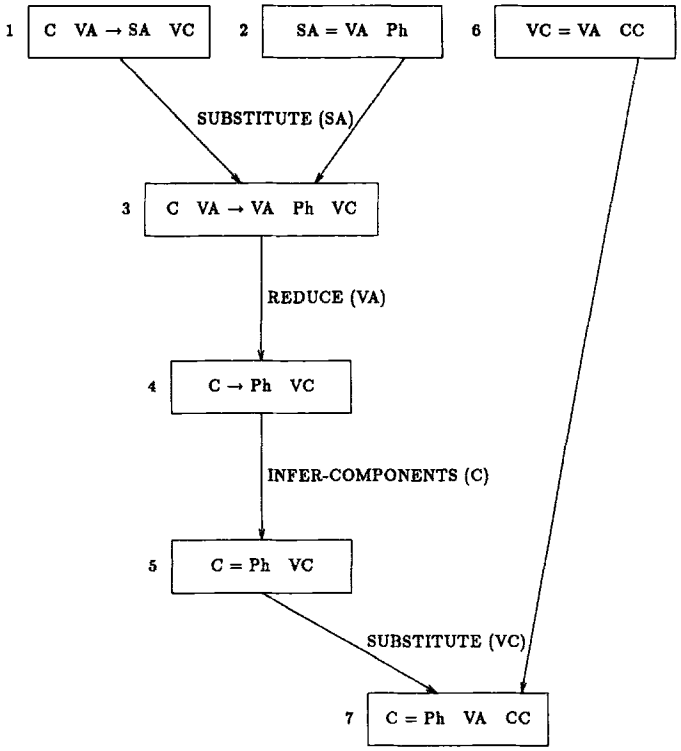

Figure 2. STAHL's revision of the copper model.

REDUCE and SUBSTITUTE rules would fire sequentially, transforming C VA → VA Ph VC into C → Ph VC before the components of VC (e.g., VA) get the chance to appear in the reaction as the result of substitution. When VA does finally reappear on the right-hand side after substitution, STAHL cannot reduce it and the system generates the incorrect copper model C = Ph VA CC.

Thus, there exists a general problem with the STAHL system: it does not know when a substance should be reduced from a reaction *due to previous applications of the REDUCE rule*. Zytkow and Simon report methods for detecting and recovering from such errors in reasoning, but we feel that these errors should be avoided in the first place. Whenever a substance is removed from any reaction by the REDUCE rule, future occurrences of that substance that appear in the reaction should also be removed. Such a mechanism forms the basis for STAHLp, and we discuss it below.

3.2 STAHLp's augmented representation

STAHLp overcomes the above problem by using an extended representation that keeps track of the substances that have been reduced from each reaction. Such a *reduced list* is stored with each reaction and componential model, and this information lets the system avoid many of the errors to which the original STAHL was subject. Let us briefly consider some examples of reduced lists before moving on to their use in the reasoning process:

calx-of-iron phlogiston ash → iron ash { }. (6)

In this case the reduced list is empty, indicating that the REDUCE rule has never been applied to reactions that led to this belief. (By definition, all input reactions or premises have empty reduced lists.) However, if STAHLp applied its REDUCE rule to reaction 5, a new belief would be generated:

clax-of-iron phlogiston → iron {ash}. (7)

The reduced list for this new reaction contains ash, the substance just removed from the reaction (5). At this point, the INFER-COMPONENTS rule would match and apply, giving the componential model:

iron = calx-of-iron phlogiston {ash}. (8)

From this we see that reduced lists are included in both reactions and models. This lets STAHLp avoid permanently adding erroneous models to memory with no possibility of corrective revision.

3.3 STAHLp's new production rules

From the above trace, we saw that STAHL's basic REDUCE rule has been slightly modified in STAHLp to take reduced lists into account. The revised REDUCE rule can be stated:

> REDUCE
> If A occurs on both sides of a reaction,
> then remove A from the reaction
> and store A in the reduced list.

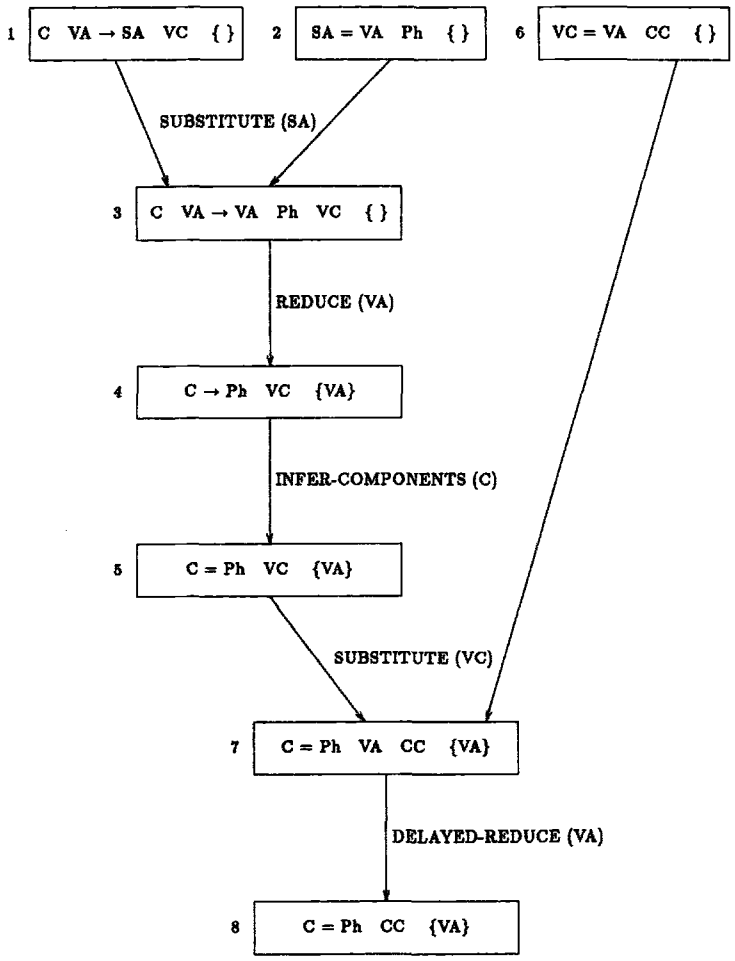

Figure 3. STAHLp's revision of the copper model.

However, this alteration is not by itself sufficient. The reduced list provides a way to remember which substances have been reduced from a reaction, but we also need some rule for removing such substances if they reappear in later (descendant) reactions. STAHLp contains just such a production:

DELAYED-REDUCE
If A occurs on either side of a reaction or model,
  and A also occurs in the reduced list,
then remove A from the reaction.

Figure 3 presents the reasoning chain that STAHLp follows to generate a componential model for copper. Although similar to the reasoning paths followed by STAHL for the same data, the presence of the reduced lists and the DELAYED-REDUCE rule makes a significant difference. Using this additional knowledge, the system infers the correct (phlogiston-based) model for copper (C = Ph CC) without difficulty.

In this trace, the system moves directly to the correct model. However, it performs equally well if the reaction VC → VA CC is added after the (incorrect) model for copper C = Ph VC is inferred. Since the DELAYED-REDUCE rule applies to models as well as to reactions, STAHLp would readily remove this model and replace it with the correct one. This example gives a flavor of the system's ability to revise its models incrementally in response to new observations.[5] However, this approach is not sufficient to eliminate all incorrect models. In the following section we examine STAHLp's methods for detecting and recovering from these more subtle errors.

4. Recovering from erroneous inferences

We have seen how STAHLp avoids one type of reasoning error that occurred in STAHL. However, Zytkow and Simon (1986) also point to three other sources of incorrect models that are unrelated to the REDUCE rule. Below we review these types of errors and present the unified approach that STAHLp uses to handle them.

4.1 A single type of error

Zytkow and Simon classified three main categories of erroneous inferences beyond those involving the REDUCE rule. First, one may generate a componential

[5] We should also note that Zytkow and Simon's system sometimes considered multiple inference paths, based on applying the inference rules in different orders. STAHLp's use of reduced lists lets it avoid this complication, so that the system need never consider more than one inference path.

model in which a substance is composed of itself, leading to infinite recursion. The pair of models A = B C and B = A D constitutes a simple example of such a circularity. Second, one may infer two incompatible componential models for the same substance; the pair of models A = B C and A = B C D provides an instance of this situation. Finally, one may infer some intermediate reaction in which one of the inputs or outputs is empty. We will call this the case of *unbalanced null reactions*.[6]

However, on closer inspection it becomes apparent that the first two error types can be viewed as instances of the third case. Let us consider some examples to clarify the mapping. Given the circularity A = B C and B = A D, one can substitute B's components into the first model. This produces the new model A = A D C, which we can then reduce to get nil = D C. This is an instance of our third error type, the unbalanced null reaction. Similarly, given the models A = B C and A = B C D (an instance of the second error type), one can substitute the first model into the second. This action produces the new 'model' B C = B C D, which can then be reduced to generate nil = D, another case of an unbalanced null reaction.

This mapping between the error types means that STAHLp can use a simple strategy for detecting errors — it need only check for unbalanced null reactions. This unified approach is more elegant than STAHL's method, which required a separate test for each type of error. However, there remains the problem of recovering from these errors. Zytkow and Simon touch on the subject when they suggest that all errors not involving the REDUCE rule are caused by 'error in the input to STAHL.' We will argue that all inconsistent models generated by STAHLp are caused by faulty premises, and that the appropriate response to such errors is to revise one or more of the input reactions that caused the difficulty. This is the problem of belief revision in chemical discovery.

4.2 Source tags

Thus far, we have depicted STAHLp's beliefs in a somewhat simplified form. In addition to the input list, output list, and reduced list, each belief includes *source tags* that indicate the premises from which each substance in that belief was derived. Source tags play the same role as assumption lists in assumption-based reasoning systems (de Kleer, 1984). When belief A is used to infer belief B, the source tags from A are propagated to B. As a result, when error recovery is required, *only a limited amount of backtracking is required.* Retaining information about the sources of reactions and models provides all the material necessary for belief revision, so there is no need to retrace one's intermediate steps. However,

[6] Balanced null reactions cause no difficulty, and in fact provide confirming evidence for the existing models.

the original premises must be remembered, because all beliefs ultimately emanate from them.[7]

Let us take a closer look at how source tags are used. Any belief in memory, whether given as a premise or inferred by STAHLp, is automatically associated with a unique number. When premises are used to infer new beliefs, these numbers are passed on, tagged to particular substances along with the side (left or right) of the premise from which the substance came. Consider the use of source tags in the five beliefs shown below, where K stands for potassium, Po for caustic-potash, H for hydrogen, and O for oxygen:[8]

1. K (1 l) { } → Po (1 r) H (1 r) { }
2. K (1 l) { } = Po (1 r) H (1 r) { }
3. Po (3 l) { } → K (3 r) O (3 r) { }
4. Po (3 l) { } → Po (1 r) H (1 r) O (3 r) { }
5. nil {Po (3 1)} → H (1 r) O (3 r) {Po (1 r)}.

Naturally, the substances contained in the initial premises (reactions 1 and 3) are tagged with the number for that reaction. However, substances in the inferred reactions and models are tagged with numbers of the initial reactions on which they are based. Also note the presence of two reduced lists — one for each side of the reaction. For instance, the caustic-potash (Po) on the left-hand side of reaction 4 has a different source than does the caustic-potash on the right-hand side. This is a common occurrence, and keeping track of the origins of the reduced list is essential to robust belief revision.

In summary, STAHLp reexamines its original 'source' reactions (i.e., premises or assumptions) during the process of error recovery. By keeping track of which premises led to which inferred beliefs, the system can generate hypotheses about which substances in the premises should be reexamined, and about how they should be modified. Now that we have considered the representational scheme that supports the process of belief revision, let us turn to the process itself.

4.3 The belief revision process

As we have seen, STAHLp incorporates a single method for detecting erroneous inferences. However, the system must still respond to these errors in some principled fashion by revising one or more of the premises (reactions) that led to the inconsistency. We can divide STAHLp's belief revision process into four

[7] Note that source tags are in the same spirit as the reduced lists, which also let STAHLp avoid backtracking through ancestral beliefs. Source tags remember which premise substances contributed to a belief, while the reduced list remembers which substances have been reduced during the inference chain leading to a belief.

[8] Actually, reaction 1 takes place in water, and thus W should appear on both sides of this reaction. We have omitted these occurrences of W for simplicity.

stages: generating alternative beliefs that would avoid the error; generating alternative premises that would lead to these beliefs; selecting the best of these revised premise reactions; and using the new reactions to infer a new (and hopefully consistent) set of beliefs.

In fact, some historical motivation exists for this approach. Chemists of the 18th century occasionally hypothesized missing substances (such as water) in order to explain conflicting experimental results. For example, Gay-Lussac and Thenard claimed that potassium consisted of caustic-potash and hydrogen, while Davy observed that caustic-potash decomposed into potassium and oxygen. To support their view, Gay-Lussac and Thenard proposed that Davy's caustic-potash had not been pure but actually contained water. As we shall see later, STAHLp has the ability to exhibit such hypothetical reasoning.

4.3.1 Generating effect-hypotheses

When STAHLp notes an unbalanced null reaction, the system responds by considering what would be required to *balance* this reaction. For example, suppose the problem reaction is nil → H O {Po}. Once detected, STAHLp deletes this belief from memory and invokes the belief revision process in order to revise the initial premises so that this error will not occur. In this case, one must ensure that H and O will not be isolated on the right-hand side of the reaction. The first step is to perform an 'inverse reduction' of all substances in the reduced lists by placing them back into the inconsistent belief. In this case, we get Po → Po H O { }; we will use the term *non-reduced reaction* to refer to such beliefs.

Next the system determines the different ways it can alter this reaction (without introducing new substances) so that the two sides balance. There are four options in this situation: (1) add H and O to the left, (2) add H to the left and delete O from the right, (3) add O to the left and delete H from the right, and (4) delete H O from the right. These are STAHLp's *effect-hypotheses* — changes to the non-reduced reaction that would have resulted if certain premises had been different. For example, if hypothesis 2 is the effect of revising the premises, then the system would infer the balanced reaction Po H → Po H instead of the inconsistent Po → Po H O.

4.3.2 Generating cause-hypotheses

The second step in belief revision involves identifying premises which, if modified, would lead to the desired effect-hypotheses and thus to a balanced reaction. These modifications take one of two forms. One can decide that a substance actually played a role in a reaction but was not observed (e.g., because it was a colorless gas). Alternatively, one can posit that one of the substances, which appearently took part, was not in fact present (e.g., that it was an illusion of some sort). We will use the term *cause-hypotheses* to refer to these possible revisions.

STAHLp uses the source tags described above to identify likely premises; they give the system immediate access to all initial reactions that led to an erroneous belief, and each such premise is a candidate for revision. This process is straight-

forward in cases involving the deletion of substances. If the system wants to eliminate some substance from a non-reduced reaction, the source tags state the premise from which it should be removed, as well as the correct side of the reaction.

For example, suppose STAHLp is working on effect-hypothesis 4 for the erroneous reaction Po → Po H O, and that this reaction includes the tags (1 r) on H and (3 r) on O. These tags tell the system that deleting H from the right-hand side of premise 1 and removing O from the right side of premise 3 will give the desired effect. As a result, the program would construct the following cause-hypothesis: *the right side of premise 1 did not have H, and the right side of premise 3 did not have O*. If put into effect, these changes will lead to the balanced reaction Po → Po.

Although cause-hypotheses involving extra substances are relatively easy to handle, hypotheses involving missing substances are more difficult. For each substance to be added in an effect-hypothesis, STAHLp must decide which premise should contain this substance to produce the desired effect. The problem is that there is no obvious source tag to employ, since the substance must be added to a premise in which it does not exist. STAHLp's solution is to use the source tags of substances that were plugged back into the empty side of the null reaction — substances that are now on the 'smaller' side of the non-reduced reaction. This is the side where substances must be added to generate a balanced reaction.

For instance, suppose the system is working on effect-hypothesis 1 from the above example. In this case, it is obvious that the left side of some premise must have really contained H and O, but STAHLp must still decide *which* premise should be revised in this manner. The source tag for the substance Po holds the answer. If the Po in the non-reduced reaction originated in the left side of premise 3, the system would hypothesize that *the Po on the left side of premise 3 actually had H and O*. If put into effect, these changes will lead to the balanced reaction Po H O → Po H O. In this example, STAHLp simply used the occurrence of Po to pinpoint the relevant premise and to determine which side to alter. This is the reason why the system plugs all reduced-list substances back into the empty side of an unbalanced null reaction — to aid in constructing cause-hypotheses that involve omissions.

In the above examples, we saw how STAHLp generates cause-hypotheses that involve removing substances from premises (1) and adding substances to premises (4). In fact, the system uses the same mechanisms to construct hybrid cause-hypotheses that involve both the addition *and* the deletion of substances, as in hypotheses 2 and 3.

4.3.3 Selecting the best hypothesis

We have seen that when STAHLp encounters an inconsistent belief, it can generate plausible explanations of the error. However, we have also seen that for any given error, there will be a number of competing hypotheses, and the system

must choose between these alternatives. When a scientist is forced to modify his theory, he is usually reluctant to alter it any more than necessary, and STAHLp incorporates a similar bias. For each initial reaction, the system keeps track of the number of models that depend on that premise. When the belief revision process suggests a number of alternative changes, STAHLp can easily compute the *cost* of making each modification — by counting the number of important beliefs (i.e., models plus the erroneous inference) that are supported by the premises about to be revised.

For instance, consider cause-hypothesis 4 from above: *the right side of premise 1 does not have H, and the right side of premise 3 does not have O.* Suppose that premise 1 happened to support seven models, while premise 3 supported two. If so, the total cost involved in revising premises 1 and 3 would be $7 + 2 + 1 = 10$; the extra cost of one represents the erroneous reaction, which both premises support. In this manner, STAHLp computes the cost associated with each cause-hypothesis and then selects the alternative with the lowest cost. This lowest-cost set of revisions will have the least impact on the existing belief structure. In cases where two hypotheses have equal costs, the system selects one of them at random.

4.3.4 Constructing a new theory

Once it has chosen the best hypothesis, STAHLp generates a new set of beliefs and models based on the revised premises. The first step is to delete the beliefs supported by each premise that will be altered by the new situation. The program retrieves these beliefs by examining their source tags; if a belief's tags include any of the modified premises, then STAHLp knows that this belief is no longer valid and removes it from memory. The second stage involves actually deleting the faulty premises and adding the new versions of these reactions to working memory. At this point, STAHLp returns to its normal inference mode, generating new reactions and componential models based on both the revised premises and those which were unchanged.

Although the new belief structure is guaranteed to avoid the error that led to revision, it may well contain new inconsistencies. In this case, STAHLp reinvokes the belief revision mechanism and further modifies its premises to eliminate the new problem. This process continues until the system generates a stable set of beliefs. Of course, additional errors may be detected as new reactions are observed, but the system has no difficulty in such cases. Thus, STAHLp can be viewed as a discovery system that carries out *incremental* revision of its theories in response to new data.

5. Examples of STAHLp's reasoning

Now that we have viewed STAHLp's mechanisms in the abstract, we can consider two examples of the system in operation. The first case involves the potash/

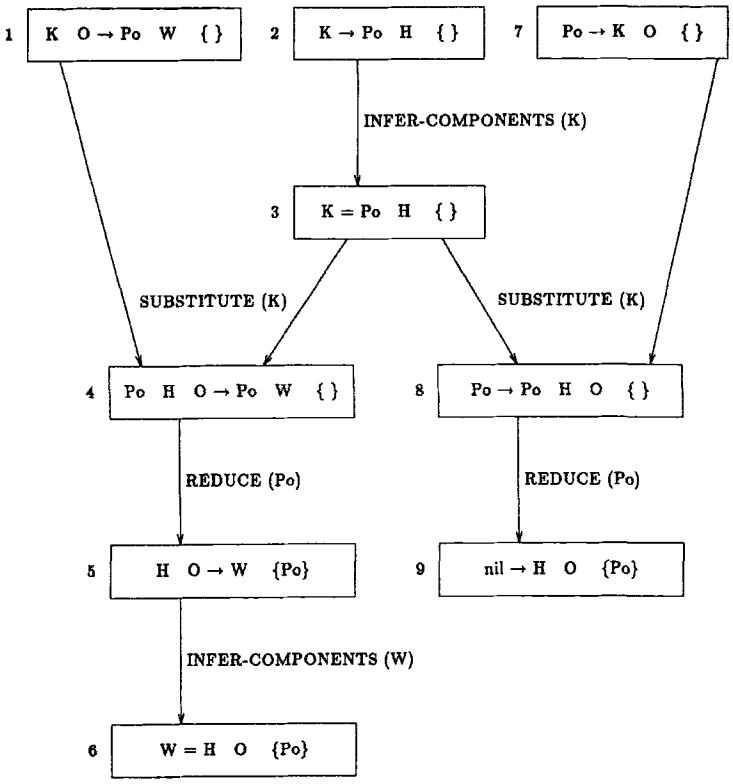

Figure 4. Inconsistent set of beliefs for potash and potassium.

potassium controversy we have already mentioned, while the second concerns the transition from phlogiston-based models to oxygen-based ones. We will see that STAHLp's belief revision process plays an important role in both cases.

5.1 STAHLp on potash and potassium

We touched earlier on the disagreement between Davy and fellow chemists Gay-Lussac and Thenard concerning caustic-potash and potassium. Now let us follow STAHLp's complete reasoning on these data. Figure 4 presents the three basic reactions[9] that were involved in the argument — 1: K O → Po W { }, 2: K → Po H { }, and 7: Po → K O { }. The figure also shows the two componential models that result from these premises. One of these (model 3) corresponds to Gay-Lussac and Thenard's belief that potassium was composed of caustic-potash

[9] We will use the following abbreviations in this example: Po, caustic-potash; K, potassium; O, oxygen; H, hydrogen; and W, water.

and hydrogen; this follows directly from reaction 2 above. The other model states that water is composed of hydrogen and oxygen; this follows from reactions 1 and 2 above.

STAHLp encounters no difficulties until it is given reaction 7, which states that caustic-potash decomposes into potassium and oxygen, which corresponds to Davy's observation. However, combining this reaction with belief 3 (that potassium contains caustic-potash and hydrogen) gives a circular definition and ultimately leads STAHLp to the unbalanced null reaction 9: nil → H O {Po}. The system immediately detects this inconsistency and invokes the belief revision process in response.

STAHLp's first step is to generate hypotheses concerning how the erroneous reaction could have been avoided. There are four balanced reactions which could have been inferred instead of the non-reduced reaction 8: Po → Po H O, had there been different initial premises: Po H O → Po H O, Po H → Po H, Po O → Po O, and Po → Po. For each case, STAHLp determines which substances must have been present or absent (and in which premises) to achieve the desired effect. This leads to four alternative effect-hypotheses:

(EH1) Po [H O] → Po H O ... left: missing H and O
(EH2) Po [H] → Po H (O) .... left: missing H; right: extra O
(EH3) Po [O] → Po (H) O .... left: missing O; right: extra H
(EH4) Po → Po (H O) ..... right: extra H and O

where brackets represent substances which should be added and parentheses indicate those which should be removed. If we ignore all substances in parentheses, each of these hypothetical reactions has balanced left-hand and right-hand sides, and can thus be transformed into a balanced null reaction using the REDUCE rule.

Now STAHLp must use these hypotheses, in conjunction with the source tags on various beliefs, to determine which premises should be modified to achieve each effect. Although we have not shown this information in the figure, reaction 8's complete representation is: Po (7 l) → Po (2 r) H (2 r) O (7 r). This tells the system that it should consider modifying premises 1 and 7, which leads to four cause-hypotheses:

(CH1) Premise 7, left side had H and O
(CH2) Premise 7, left side had H
Premise 7, right side did not have O
(CH3) Premise 7, left side had O
Premise 2, right side did not have H
(CH4) Premise 7, right side did not have O
Premise 2, right side did not have H

which correspond to the four effect-hypotheses shown above. Note that STAHLp

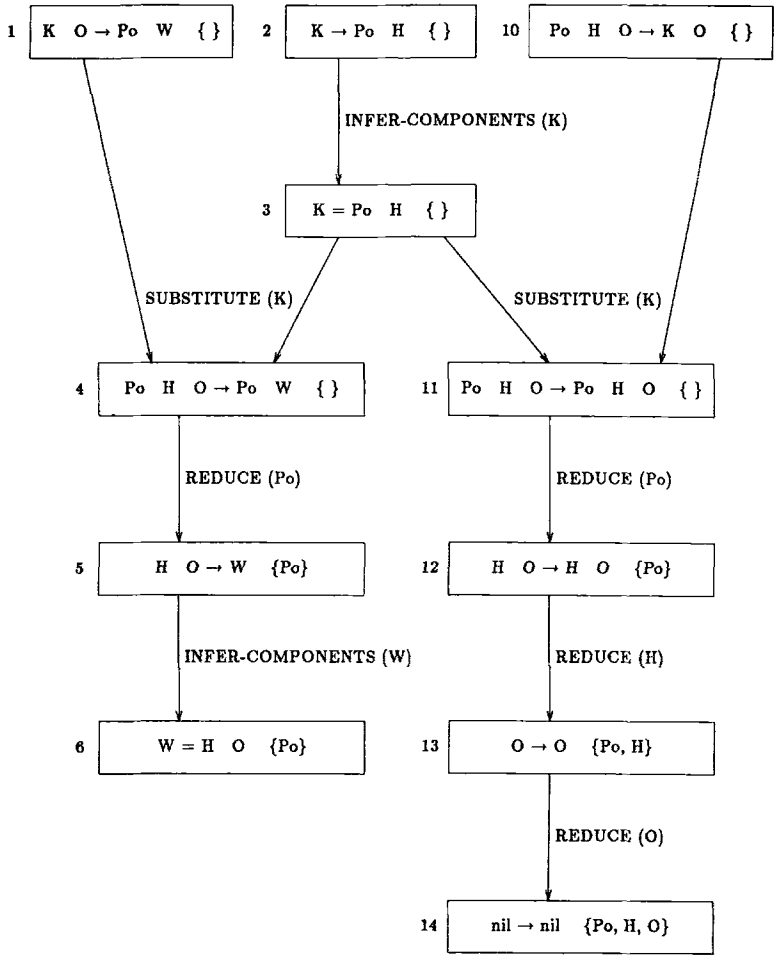

Figure 5. Consistent set of models after belief revision.

exhibits dependency-directed reasoning (Doyle, 1979) by not constructing hypotheses involving premise 1, since it was not involved in inferring the unbalanced reaction.

Now that STAHLp has generated specific revisions, it must select among the four competitors. To do this, it computes the 'cost' of carrying out each change. Observe that premise 2 supports three important beliefs, while premise 7 supports only one (counting the models plus the unbalanced reaction in our measure of importance). This leads to the following costs for each hypothesis:

(CH1) would change belief 9; cost = 1
(CH2) would change belief 9; cost = 1
(CH3) would change beliefs 3, 6, and 9; cost = 3
(CH4) would change beliefs 3, 6, and 9; cost = 3.

Thus, hypotheses CH1 and CH2 tie for the lowest cost, and the system arbitrarily selects CH1 as the best hypothesis. In other words, STAHLp modifies Davy's reported reaction (belief 7) by adding both H and O to its right-hand side, giving the new premise 10: Po H O → K O. This is very similar to Gay-Lussac and Thenard's claim that Davy's caustic-potash actually contained some water, and that this was the source of his odd results.[10] Once this revision has been implemented, STAHLp removes all beliefs that were based on the old reaction and then proceeds to make inferences based on the new premise. Figure 5 shows the results of this process. Note that the end result is a balanced null reaction (14), indicating that the new premise is consistent with the other reactions in memory.

5.2 The shift from phlogiston to oxygen

We now turn to another example from 18th century chemistry that has a similar structure to the potassium/potash case, but which historically had a much greater impact. The prevailing theory of this period stated that, during combustion, a substance called phlogiston left the consumed body and entered the surrounding air. The phlogiston chemists hypothesized that a similar process occurred during the calcination (rusting) of metals. One such reaction involved the transformation of mercury, upon exposure to air, into the substance calx-of-mercury. We present this reaction as premise 1 in Figure 6, along with the componential model that follows directly — that mercury is composed of calx-of-mercury and phlogiston (belief 2).

However, in the 1770s, Joseph Priestley produced another reaction that introduced difficulties. He found that, upon heating, calx-of-mercury generated mercury and a colorless gas (which we now call *oxygen*). We show this reaction as premise 3 in the figure. The circularity is clear upon inspection; calx-of-mercury contains mercury, but mercury also contains calx-of-mercury. Zytkow and Simon's (1986) STAHL system responds to such circularities by renaming one of the substances in the componential models that are inferred; thus, one occurrence of calx-of-mercury might be replaced by 'calx-of-mercury-proper.'

Although this strategy avoids the infinite recursion, we believe that STAHLp's response is both more meaningful and a better model of historical reasoning. When presented with reactions 1 and 3 in Figure 6, the program notes the circularity in terms of an unbalanced null reaction (belief 5). The system then invokes its belief revision process to identify and alter one of these premises. There are four ways to balance the offending null reaction — CM Ph O → CM Ph O, CM Ph → CM Ph, CM O → CM O, and CM → CM — each leading to a different effect-hypothesis that would generate the balanced reaction:

[10] We now know that Davy's model was the correct one, with potassium the element and caustic-potash the compound. However, Gay-Lussac and Thenard's reasoning was quite plausible, given the data available at the time, and it is this plausible reasoning that we intend STAHLp to model.

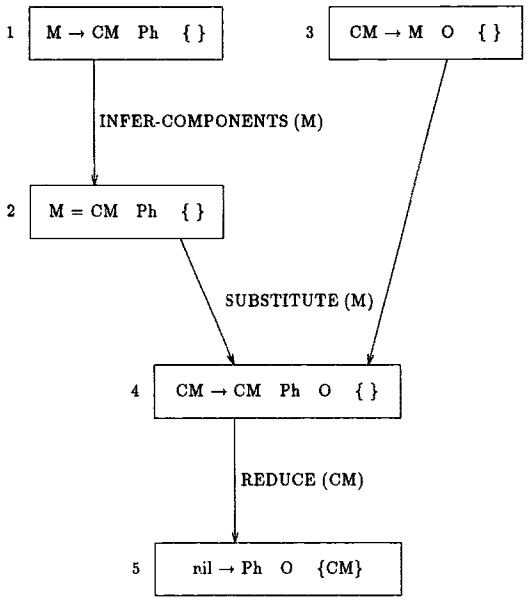

Figure 6. Inconsistent set of beliefs for mercury and calx-of-mercury.

(EH1) CM [Ph O] → CM Ph O .... left: missing Ph and O
(EH2) CM [Ph] → CM Ph (O) .... left: missing Ph; right: extra O
(EH3) CM [O] → CM (Ph) O ..... left: missing O; right: extra Ph
(EH4) CM → CM (Ph O) ..... right: extra Ph and O

and these are used in turn to produce four cause-hypotheses, one for each of the effect-hypotheses:

(CH1) Belief 3, left side had Ph and O
(CH2) Belief 3, left side had Ph
Belief 3, right side did not have O
(CH3) Belief 1, left side had O
Belief 1, right side did not have Ph
(CH4) Belief 3, right side did not have O
Belief 1, right side did not have Ph.

Next, STAHLp computes the cost of each hypothesis in terms of the number of beliefs that would be revised in each case. CH3 and CH4 each affect two important beliefs (the model plus the erroneous reaction), but the first two hypotheses tie with a score of 1, so one would be selected at random. Let us consider

the historical significance of two of these hypotheses.[11]

The second hypothesis (CH2) states that the reaction CM → M O was actually CM Ph → M, with the oxygen being 'imagined' and the phlogiston being unobserved. In contrast, hypothesis CH3 states that the reaction M → CM Ph was actually M O → CM, with the phlogiston being 'imagined' and the oxygen being unobserved. One of the phlogiston chemists might well take such a stance, refusing to believe the evidence against a theory that had worked so well. We can imagine that in a more typical historical scenario, where other reactions consistent with the phlogiston theory existed (and which were based on reaction 1), the cost of CH3 and CH4 would be even higher; thus, rejecting such hypotheses in favor of another (like CH1 or CH2), which requires fewer revisions, would be even more plausible. Thus, we believe STAHLp provides a very plausible model of 'normal science,' in which occasional anomalies are ignored in order to save an existing theory.[12]

On the other hand, one of the early proponents of the oxygen theory might well take the opposite view, rejecting the existence of phlogiston and favoring CH3 over CH2. Giving STAHLp other reactions consistent with the oxygen theory (and based on reaction 3) might increase the cost of CH2 so that it would be rejected. However, this account does not seem very plausible historically. At the time Lavoisier and his followers rejected the phlogiston theory, many of the chemical community's beliefs were linked to phlogiston, and the 'cost' of replacing any single phlogiston-related reaction with an oxygen-based interpretation would have been quite high. Even though STAHLp shows how Lavoisier might have generated his basic hypothesis, it provides no convincing model of the decision to follow this lead. Our system does not account for 'revolutionary science.'

An improved model of the shift from phlogiston to oxygen would require an improved cost measure that accumulates evidence over time. Rather than selecting one hypothesis and rejecting the others, the system would store each hypothesis for future reference. If later inconsistences lead to the same revision being proposed, the evidence count for that hypothesis would be incremented. When the evidence for a given hypothesis exceeded the evidence for an existing reaction, the program would reject the current premise and implement the hypothesis at that time. In our example, one might begin to suspect something was

[11] Hypothesis CH1 proposes a compromise of sorts, replacing the reaction CM → M O with CM Ph O → M O. This allows both phlogiston and oxygen to exist, but states that both substances went unobserved in the reaction. In fact, Priestley suggested a similar explanation of the calx-of-mercury experiment.

[12] This brings out another interesting feature of STAHLp. As long as the input reactions are correct, the system generates the same models independent of the presentation order. However, when inconsistent data are provided, order effects become very important. We believe this is desirable in a model of historical discovery.

amiss with the phlogiston theory and gradually garner evidence in favor of competing views (like the oxygen-based CH3). Eventually, one might accumulate enough support to justify rejecting even long-standing premises upon which many beliefs rest. We believe this scheme provides a more plausible account of Lavoisier's decision, and we plan to implement it in future versions of STAHLp.

Also, the proponents of both oxygen and phlogiston formulated general qualitative laws that summarized the types of reactions that could occur. For instance, Lavoisier proposed generic patterns for combustion and calcination:

combustible + oxygen-gas → oxide + caloric
metal + oxygen-gas → metallic-oxide + caloric

in which the terms *combustible*, *oxide*, *metal*, and *metallic-oxide* represented entire classes of substances. A new reaction was plausible to the extent that it obeyed one of these general reaction schemas. We would like STAHLp to incorporate such reasoning in its evaluation of hypotheses, and we would also like our system to generate such qualitative laws on its own initiative. In fact, Langley, Simon, Bradshaw, and Zytkow (1986) and Jones (1986) have described GLAUBER, a system that discovers qualitative empirical laws of just this form. We also plan to incorporate methods from GLAUBER in future versions of STAHLp.

Finally, some cause-hypotheses seem more plausible that others on the basis of world knowledge. For instance, it seems quite plausible that a chemical prepared by drying from solution might retain undesired water, as Gay-Lussac and Thenard claimed about Davy's caustic potash. However, other hypothesized omissions or commissions are much less plausible, and STAHLp in its current form has no way of distinguishing them. Future versions of the system should use additional knowledge during the belief revision process. Hopefully, we can represent this knowledge as additional premises and thus require minimal changes in the structure of the system.

6. Generality of the system

Generality is an important measure of any machine learning system's success, so let us consider how well STAHLp performs along this dimension. Below we summarize the classes of reactions the program handles, dividing them into groups that involve different modes of reasoning. We also consider a set of reactions that gave problems to Zytkow and Simon's system but which STAHLp handles without difficulty. Finally, we show that, with minor modifications, our system can replicate an entirely different kind of chemical reasoning.

6.1 Simple reactions

Typically, chemical reasoning of the 18th century was based on correct reactions and so provides little challenge to our system's heuristics. Table 1 presents four

Table 1. Simple reactions on which STAHL and STAHLp agree.

Inputs	STAHL/STAHLp outputs
(a) Charcoal air → phlogiston ash air	Charcoal = phlogiston ash
Calx-of-iron charcoal air → iron ash air	Iron = phlogiston calx-of-iron
Charcoal litharge → lead ash	Lead = phlogiston litharge
Vitriolic-acid potash → vitriolic-tartar	Vitriolic-tartar = vitriolic-acid potash
Sulfur potash → liver-of-sulfur	Liver-of-sulfur =
Vitriolic-tartar charcoal → liver-of-sulfur	vitriolic-acid potash phlogiston ash
Copper vitriolic-acid →	Sulfur = vitriolic-acid phlogiston ash
sulfurous-acid vitriol-of-copper	Copper = phlogiston vitriol-of-copper
Sulfurous-acid → vitriolic-acid phlogiston	Sulfurous-acid = vitriolic-acid phlogiston
(b) Potassium water →	Green-solid = caustic-potash ammonia
caustic-potash hydrogen water	Water = hydrogen oxygen
Caustic-potash water → potassium oxygen	Potassium = caustic-potash hydrogen
Green-solid water →	
caustic-potash ammonia water	
Potassium ammonia → hydrogen green-solid	
(c) Lime → quicklime fixed-air	Gypsum = quicklime vitriolic-acid
Quicklime magnesia-alba →	Vitriolic-tartar = vitriolic-acid caustic-potash
lime calcined-magnesia	Epsom-salt = calcined-magnesia vitriolic-acid
Quicklime salt-of-tartar → lime caustic-potash	Salt-of-tartar = fixed-air caustic-potash
Lime vitriolic-acid → gypsum fixed-air	Magnesia-alba = fixed-air calcined-magnesia
Magnesia-alba vitriolic-acid →	Lime = quicklime fixed-air
epsom-salt fixed-air	
Quicklime vitriolic-acid → gypsum	
Caustic-potash epsom-salt →	
calcined-magnesia vitriolic-tartar	
Calcined-magnesia vitriolic-acid → epsom-salt	
(d) Oxygen-gas = oxygen-principle caloric	Oxygen-gas = oxygen-principle caloric
Calx-of-lead caloric → lead oxygen-gas	Calx-of-lead = lead oxygen-principle
Calx-of-lead caloric → lead oxygen-gas	Fixed-air = oxygen-principle charcoal caloric
Calx-of-lead charcoal caloric → lead fixed-air	Water = oxygen-principle hydrogen-gas
Water charcoal caloric → hydrogen-gas fixed-air	Oxide-of-iron = oxygen-principle iron caloric
Water iron caloric → hydrogen-gas oxide-of-iron	
Charcoal oxygen-gas → fixed-air caloric	

sets of such reactions for which STAHL and STAHLp generate the same componential models. No belief revision is required, since this occurs only when erroneous inputs are present. However, this does not mean that the models have no historical interest.

Set (a) in the table consists of eight reactions that support the phlogiston theory. The reactions contain a variety of substances, including charcoal, iron, copper, lead, sulfur, and their associated calxes and vitriols. Given these reactions, STAHLp generates eight componential models; historically, some of these were later revised as substances once thought to be compounds (e.g., iron) were found to be elements, and vice versa. Zytkow and Simon report runs with their STAHL system on subsets of these reactions, in addition to runs on the entire set;

STAHLp replicates their results in this mode as well.

The reactions labeled (b) in Table 1 summarize the data from which Gay-Lussac and Thenard (1808, 1810) reached their conclusions about caustic-potash and potassium. The data here are more complicated than those in our earlier example on this topic, but we are not concerned here with the belief revision process. In any case, both STAHL and STAHLp reach the conclusion that potassium is composed of the 'elements' caustic-potash and hydrogen, just as did the French chemists.

The third reaction set (c) contains data reported by Black (1756) in his work on fixed air (carbon dioxide) and alkaline substances. From these eight reactions, the chemist inferred six componential models that specified structures for lime, gypsum, epsom-salt, magnesia-alba, vitriolic-tartar, and salt-of-tartar. Given the same premises, both STAHL and STAHLp arrive at the same componential models as did Black.

Reaction set (d) specifies six reactions that Lavoisier used to justify models based on his oxygen theory of combustion. Actually, the first reaction is more a theoretical statement than an observed reaction, but Zytkow and Simon note that, without this input, their STAHL system cannot generate componential models from the remaining five reactions. Given all six premises, both STAHL and STAHLp arrive at models equivalent to those Lavoisier proposed in the late 18th century. However, we will see later that STAHLp can make some inferences even without the additional input.

6.2 Identification of substances

Zytkow and Simon (1986) point out that the early chemists sometimes decided, on the basis of analogous models, that two apparently different substances were in fact the same chemical. Table 2 presents two sets of reactions from which STAHL drew such conclusions. Their system employed two additional heuristics in making such identifications, one for collapsing substances with the same componential model and another for combining substances that had been inferred as components of the same compound. We have included similar production rules in the STAHLp system.

Set (a) in Table 2 summarizes additional steps in Black's reasoning about alkalines and fixed air. In this case, the system is provided with the componential models from the earlier run along with two new reactions. Taken together, these beliefs lead both STAHL and STAHLp to decide that lime, calcite, and chalk are actually the same substance, and that this chemical is composed of quicklime and fixed-air.

Set (b) contains the reactions used by Berthollet to infer that chlorine is composed of muriatic-acid and oxygen. Along the way, both systems also conclude that chlorine and oxymuriatic-acid are the same substance. This reaction set is interesting because of its redundancy. The first two premises lead one to the same

Table 2. Reactions involving identification of substances.

Inputs	STAHL/STAHLp outputs
(a) Calcite vitriolic-acid → gypsum fixed-air Chalk vitriolic-acid → gypsum fixed-air Gypsum = quicklime vitriolic-acid Vitriolic-tartar = vitriolic-acid caustic-potash Epsom-salt = calcined-magnesia vitriolic-acid Salt-of-tartar = fixed-air caustic-potash Magnesia-alba = fixed-air calcined-magnesia Lime = quicklime fixed-air	Lime-calcite-chalk = quicklime fixed-air Gypsum = quicklime vitriolic-acid Vitriolic-tartar = vitriolic-acid caustic-potash Epsom-salt = calcined-magnesia vitriolic-acid Salt-of-tartar = fixed-air caustic-potash Magnesia-alba = fixed-air calcined-magnesia
(b) Chlorine water → oxymuriatic-acid water Oxymuriatic-acid water → muriatic-acid oxygen water Black-manganese → calcined-manganese oxygen Black-manganese muriatic-acid water → salt-of-manganese chlorine water Calcined-magnesia muriatic-acid water → salt-of-manganese water	Chlorine-oxymuriatic-acid = muriatic-acid oxygen

models as the last three reactions, providing a check on the correctness of the inferences.

6.3 Differences between STAHL and STAHLp

We have seen that STAHLp differs from its predecessor in two ways — in its ability to recall which substances have been reduced from a reaction, and in its ability to revise input reactions when these lead to inconsistencies. Table 3 presents reactions in which these differences lead the systems to produce different componential models.

The first set of reactions (a) summarizes the copper example used to introduce the notion of a reduced list. In this case, Zytkow and Simon's system actually arrives at an incorrect componential model, including vitriolic-acid in the model for copper where it does not belong. In contrast, STAHLp generates a simpler model that is correct, given the input premises.

The reactions involving mercury and calx-of-mercury constitute the second set (b) in the table. These data lead STAHL to introduce a 'conceptual distinction' between the two occurrences of calx-of-mercury in order to avoid the circular definition. In contrast, STAHLp invokes its belief revision process and modifies the more recent reaction, deciding to retain its phlogiston-based model.

The third set of reactions (c) in the Table relates to a completely different issue. We saw in Table 1 (d) that, in order to replicate Lavoisier's reasoning, Zytkow and Simon's system required the unobserved premise oxygen-gas = oxygen-

Table 3. Reactions on which STAHL and STAHLp disagree.

Inputs	STAHL outputs	STAHLp outputs
(a) Vitriol-of-copper → vitriolic-acid calx-of-copper Copper → phlogiston vitriol-of-copper Sulfurous-acid → vitriolic-acid phlogiston	Vitriol-of-copper = vitriolic-acid calx-of-copper Copper = phlogiston vitriolic-acid calx-of-copper Sulfurous-acid = vitriolic-acid phlogiston	Vitriol-of-copper = vitriolic-acid calx-of-copper Copper = phlogiston calx-of-copper Sulfurous-acid = vitriolic-acid phlogiston
(b) Mercury → phlogiston calx-of-mercury Calx-of-mercury → mercury oxygen	Mercury = phlogiston calx-of-mercury-proper Calx-of-mercury = mercury oxygen	Mercury = phlogiston calx-of-mercury
(c) Charcoal oxygen-gas → fixed-air caloric Calx-of-lead caloric → lead oxygen-gas Water iron caloric → hydrogen-gas oxide-of-iron Water charcoal caloric → hydrogen-gas fixed-air Calx-of-lead charcoal caloric → lead fixed-air	[none]	Oxide-of-iron = oxygen-gas iron Fixed-air = charcoal oxygen-gas Water caloric → oxygen-gas hydrogen-gas Calx-of-lead caloric → oxygen-gas lead

principle caloric in addition to the five observed reactions. This seems historically plausible, since Lavoisier invoked such a belief to justify his oxygen-based models, and we saw that STAHLp could replicate this reasoning as well.

However, our system also includes an additional method that increases its ability to infer componential models. Given a situation in which no models are proposed by the rules we have discussed, STAHLp invokes an additional production rule that lets it 'subtract' one reaction or model from another. This produces a new reaction which, though it may never occur in nature, may provide the material for STAHLp's basic inference methods to operate upon.

In the Lavoisier example, the system 'subtracts' the reactions calx-of-lead caloric → lead oxygen-gas and calx-of-lead charcoal caloric → lead fixed-air. This action generates the new reaction fixed-air → oxygen-gas charcoal, which leads directly to a componential model. This leads in turn to other inferences, until the system arrives at the models shown in Table 3 (c). Although we will not claim that the early chemists used such an inference rule, it nevertheless constitutes an interesting addition to STAHLp's repertoire.

6.4 Constructing molecular models

As a final example of STAHLp's generality, let us consider its application to the formulation of molecular models. Between 1805 and 1815, chemists like Dalton and Avogadro began to propose simple structural models for compounds such as water and ammonia. However, they used different criteria and thus arrived at competing models for the same substances. Dalton relied on his rule of greatest simplicity, which led him to the model W = h o for water and A = n h for ammonia. In contrast, Avogadro used Gay-Lussac's law of combining volumes to constrain his models, arriving at W = h h o for water and A = n h h h for ammonia.

Langley et al. (1986) have described DALTON, a discovery system that replicates the reasoning of both chemists. The system accepts as input both reactions and componential models (like those generated by STAHLp). As output, the program generates molecular models of both elements and compounds, like those shown above. Thus, it would produce models for hydrogen, oxygen, and nitrogen as well as for water and ammonia. In the Dalton version, these would be H = h, O = o, and N = n, where uppercase represents molecules and lowercase represents atoms. In the Avogadro variant, the models would be H = h h, O = o o, and N = n n. The first set contains monatomic models, while the second posits diatomic models.

The DALTON system carries out a depth-first search through the space of molecular models, using known reactions and conservation of particles to constrain this search. In modeling Dalton's reasoning, the program assumes that reactions involve only single molecules, and this leads to consistent models for water and ammonia with no backtracking. In modeling Avogadro's reasoning, the system employs knowledge of the relative volumes involved in the water and ammonia reactions to determine the number of molecules in those reactions. This causes some backtracking when determining the number of atoms in each molecule, but eventually the system arrives at the (correct) diatomic models shown above.

DALTON was designed specifically to simulate the formation of molecular models, but we have found that STAHLp can replicate this process with only minor modifications. In particular, removing the rule for delayed reduction lets the revised system (call it STAHLp′) construct molecular models in which the same substance occurs multiple times. However, inconsistencies can still arise when the program infers unbalanced reactions, as well as reactions involving only atoms, such as h → n. In these cases, STAHLp′ invokes its belief revision process and modifies one or more of its premises. This simulates DALTON's backtracking through the space of molecular models.

Let us consider the system's behavior on the water and ammonia reactions. Suppose we give STAHLp′ initial monatomic molecular models for hydrogen and oxygen (H = h and O = o), along with the simplest possible model for water

(W = h o). No new beliefs can be inferred from this information, but this changes when we give the water reaction to the system: H H O → W W.[13] Substitution leads to the reaction h h o → h h o o, and this in turn produces the unbalanced null reaction nil → o. STAHLp′ invokes the belief revision process in response, generating O = o o as the revised model for oxygen. This model is consistent with the premises H = h and W = h o and thus constitutes an acceptable summary of the data, even though it differs from the modern view.

However, when the system encounters the initial models for nitrogen (N = n) and ammonia (A = n h), along with the reaction H H H N → A A, further revisions become necessary. Substitution leads to the reaction h h h n → h h n n, which in turn produces h → n. Belief revision generates the revised premises N = n n and A = n h h, which eliminate the inconsistent reaction. However, these lead to the reaction h h h n n → h h h h n n and thus to nil → h, another inconsistency. This time belief revision modifies the model for hydrogen, giving H = h h.

Progress has occurred, but STAHLp′ has still not converged on a consistent set of premises. At this point, substitution produces the reaction h h h h h h n n → h h h h n n, which then leads to h h → nil. In response, belief revision alters the molecular model for ammonia, giving A = n h h h. These revisions produce a consistent (and correct) model for the ammonia reaction, but the diatomic hydrogen model introduces problems for the water reaction, giving h h h h o o → h h o o. This generates the unbalanced reaction h h → nil, and in this case belief revision alters the model for water, giving the modern-day model W = h h o. At this point, the system has generated a consistent set of premises and halts its cycle of revision and testing.

In summary, removing the delayed reduction rule lets STAHLp replicate Avogadro's reasoning about the water and ammonia reactions, without need for the additional search control used by Langley et al.'s DALTON. This suggests that the belief revision methods we have employed are more general than the inference rules themselves, and that we may model scientific reasoning in other domains by adding or removing heuristics to represent the constraints relevant for each domain. In some cases, these constraints may be very strong and belief revision may be invoked only rarely. In other cases, the constraints may be weaker and the belief revision process may be invoked many times — effectively producing heuristic search, as we saw in the DALTON example above.

7. Summary

In this paper we have described STAHLp, a system that formulates componential models of chemical substances from input reactions. The program has many

[13] Multiple occurrences of a symbol represent the number of molecules involved in the reaction, based on Gay-Lussac's law of combining volumes.

similarities to Zytkow and Simon's (1986) STAHL, but our system also includes a number of improved features. One of these is the ability to recall when a substance has been reduced from a given reaction, and to use this knowledge to avoid simple reasoning errors. More importantly, STAHLp incorporates a unified mechanism for detecting more subtle reasoning errors and for recovering from these mistakes.

This belief revision process rests on two assumptions: that all errors ultimately lead to unbalanced null reactions, and that all errors are caused by incorrect input reactions in which substances were omitted or extra substances were included. The basic process is similar to de Kleer's (1984) assumption-based method in that STAHLp retains information about the premises used to infer its beliefs, and uses this information to identify faulty input reactions. This method gives STAHLp the ability to revise its chemical theories incrementally in response to new observations, and lets the system model two cases from 18th century chemistry in which disagreement arose over the interpretation of reactions.

STAHLp has shown its generality by replicating historical reasoning on a number of different sets of reactions. In one case, the system provides a partial account of the transition from the phlogiston theory to the oxygen theory of Lavoisier. In another, minor modifications let the program model the formation of simple molecular models. Despite these successes, a number of extensions suggest themselves, including improved measures of evidence for use in belief revision and the formulation of general qualitative laws that summarize specific reactions. Nevertheless, we feel that STAHLp provides an illuminating account of early chemical reasoning, and that it suggests a promising paradigm for integrating research on machine discovery and belief revision.

Acknowledgments

We would like to thank Jan Zytkow, Chris Truxaw, Bernd Nordhausen, Randy Jones, and Dennis Kibler for their valuable comments on the ideas presented in this paper. This work was supported by contract N00014–84–K–0345 from the Information Sciences Division, Office of Naval Research.

References

Black, J. (1756). Experiments upon magnesia alba, quicklime, and some other alkaline substances. In *Essays and observations, physical and literary*. Edinburgh.

de Kleer, J. (1984). Choices without backtracking. In *Proceedings of the Fourth National Conference on Artificial Intelligence* (pp. 79–85). Austin, Tx: Morgan-Kaufmann.

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.

Jones, R. (1986). Generating predictions to aid the scientific discovery process. In *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 513–517). Philadelphia: Morgan-Kaufmann.

Langley, P., Ohlsson, S., Thibadeau, R., & Walter, R. (1984). Cognitive architectures and principles of behavior. In *Proceedings of the Sixth Conference of the Cognitive Science Society* (pp. 244–247). Boulder, Co.

Langley, P., Simon, H.A., Bradshaw, G.L., and Zytkow, J.M. (1986). *Scientific discovery: A computational account of the creative processes*. Cambridge, MA: MIT Press.

Stillman, J.M. (1960). *The story of alchemy and early chemistry*. New York: Dover Publications.

Zytkow, J.M., & Simon, H.A. (1986). A theory of historical discovery: The construction of componential models. *Machine Learning*, *1*, 107–136.