MULTIPLE THEORIES IN SCIENTIFIC DISCOVERY[1]

Donald Rose
Department of Information & Computer Science
University of California, Irvine CA 92717 USA

INTRODUCTION

In this paper we describe enhancements being made to REVOLVER, a program that formulates componential models of objects (e.g., physical substances) and replicates historical discoveries from domains such as chemistry and physics. The program inputs reactions relating groups of substances and uses heuristics to transform these premises into models. If premises lead to inconsistent beliefs, the system searches the space of revised premises in order to resolve the errors. The system was originally designed to process only one theory at a time (i.e., to keep only one theory in its database), using hill climbing as its search strategy; once a belief was revised in response to an error, the old belief was deleted.

The enhancements described in this paper deal mainly with extending the REVOLVER framework to handle multiple theories. First, the enhanced system explicitly represents assertions as well as an agent's belief in assertions (meta-assertions). Second, inferencing is performed only on meta-assertions, allowing separate theories to form for multiple agents. By explicitly separating belief in an assertion from the assertion itself, independent theories can coexist for several agents, even though the inference rules used by each can be the same. Third, meta-assertions can indicate how strongly an assertion is believed; this degree of belief could then be used to perform inferencing on strongly-held beliefs first, while it can also be used to bias the system such that they are revised last. The other main enhancement involves explicit representation of unknown objects in system beliefs, whereas the original program only dealt with objects already known to be part of its beliefs.

In the following pages we describe REVOLVER's basic inference and revision processes, then discuss the proposed enhancements to the program. An example from the history of science is then presented which illustrates how the new enhancements can be utilized. Finally, related work plus ideas for future improvements are discussed.

THE REVOLVER SYSTEM

Reactions and *models* are the two kinds of beliefs used by the program. Reactions represent relations between objects and are given as input premises. A premise might represent the inputs and outputs of a chemical reaction. An example from 18th century chemistry would be the observation that potassium and oxygen react to form caustic-potash and water, or (1) K O $\rightarrow$ P W. Another observation from that era was that potassium reacts to form caustic-potash and hydrogen, or (2) K W $\rightarrow$ P H W. Given premises such as these, REVOLVER tries to infer new models of substances by using a set of general heuristics. Since W appears on both sides of (2), it is reduced from the reaction, leaving (3) K $\rightarrow$ P H. Whenever a substance is alone on one side of a reaction, the system infers that its components are present on the opposite side. Thus REVOLVER now infers from (3) the model (4) K = P H. The program then substitutes K's components from (4) into (1) to get

[1]This research was supported by Contract N00014-84-K-0345 from the Information Sciences Division, Office of Naval Research. I would like to thank Pat Langley, Paul Thagard, Randy Jones and Bernd Nordhausen for discussions that helped develop and refine the ideas in this paper.

(5) P H O → P W. Next, P is reduced from both sides of (5), leaving (6) H O → W. The final inference is the model (7) W = H O.

In the above example, the premises were consistent; REVOLVER reached a quiescent state without inferring any erroneous beliefs. However, sometimes the premises given to REVOLVER lead to reactions having either no inputs or no outputs. In order to restore premises to consistency, REVOLVER invokes its belief revision process. The program finds the premises that led to the inconsistency, and considers revisions to them that will lead REVOLVER closer to a consistent set of beliefs. After revising a premise, the system continues making new inferences and, if it detects new inconsistencies, again revises premises. This cycle continues until no more inferences can be made and no inconsistencies exist.

Continuing our example, suppose the system receives new premise (8) P → K O. Substituting for K leads to (9) P → P H O and reducing P yields inconsistent reaction (10) nil → H O. Invoking belief revision, REVOLVER identifies premises (2) and (8) as the sources of the error, and must decide how to revise them in order to resolve the inconsistency.

REVOLVER now tries to eliminate each substance in the inconsistency, one at a time. Hence, in our example, the system proposes six candidate revisions:

R1: add H to (2)'s inputs; R2: add O to (2)'s inputs; R3: delete H from (2)'s outputs;
R4: add H to (8)'s inputs; R5: add O to (8)'s inputs; R6: delete O from (8)'s outputs.

Implementing any of these revisions would remove *one* substance from the inconsistency after REVOLVER has made the change selected, deleted impacted beliefs, and restarted its basic inference process.

Only one of the candidate revisions in REVOLVER is actually carried out. Thus, selecting the best revision(s) is especially important; the program uses an *evaluation function* to make the selection. REVOLVER scores the premises considered for revision along several criteria, multiplies each score by a weight (indicating the priority given to each criterion), sums the weighted scores, and revises the premise(s) having the lowest total score. Since the program does not retain alternate revised premises after each revision step, it is a hill-climbing system, relying on its evaluation function to intelligently guide search towards consistent beliefs. The criteria for selecting the best revision are discussed elsewhere (Rose, 1988; Rose & Langley, 1988).

Continuing our example, let us assume that R5 is selected. This results in revised premise (11) P O → K O. Substituting P and H for K in (11) and reducing P and O yields (14) nil → H. Now REVOLVER proposes revisions again:

R7: add H to (2)'s inputs; R8: delete H from (2)'s outputs; R9: add H to (11)'s inputs.

Now let us suppose R9 is selected. This results in (15) P H O → K O. After making this revision and restarting the basic inference process, substitution leads to P H O → P H O, which after reductions leads to nil → nil, indicating consistency has been reached.

ENHANCING REVOLVER

While the example just seen is a reasonable model of how 18th century chemists reasoned about reactions, there were a number of drawbacks in the REVOLVER framework as presented above. The first main drawback was that it did not reason about multiple theories. Constructing inferences from premises asserted (or recognized) in one order can sometimes lead to a different theory than when the same premises are asserted in some different order. However, the program could only

investigate one of these possibilities. In addition, multiple theories could not coexist in its database. The second main drawback of the system is that it did not separate belief in assertions from the assertions themselves. The third main drawback of REVOLVER is that it did not reason about degrees of belief. That is, one assertion was never believed more than any another.

This paper discusses techniques being developed to enhance the system so that each of these areas are addressed. First, the enhanced system explicitly represents assertions as well as an agent's belief in assertions (meta-assertions). Second, inferencing is performed only on meta-assertions, allowing separate theories to form for multiple agents. Third, meta-assertions indicate how strongly an assertion is believed; this degree of belief can then be used to perform inferencing on strongly-held beliefs first, while it can also be used to bias the system such that they are revised last.

Inferring Multiple Theories

The new version of REVOLVER not only represents assertions, but also represents agents' belief in those assertions. I call the latter *meta-assertions*. For example, an assertion might be the observation (2) K W → P H W. The meta-assertion that agent1 believes (2) would be agent1(K W → P H W). The separation of belief in assertions from the assertions themselves enables multiple theories to coexist in REVOLVER's database. Theories can be created for each agent, each theory being independent of the other.[2] These mutually independent theories result because inferencing in the enhanced REVOLVER is done on meta-assertions only. In other words, beliefs must be recognized by an agent before a new inference can be made; the new inference is then automatically assumed to be held by that same agent.

This enhancement means that even if two agents use the same premises and the same inference rules, different theories may result. This is due to the effects of assertion ordering on the inference process. The order in which agents recognize premises influences the order in which inferences are made, and since different orderings can lead to different theories, it is possible for agents to hold different theories even if their initial premises are the same. In the original REVOLVER, order was never an issue; the most recently asserted premises were always processed immediately, and any new premises were processed as they arrived. The new ability to explicitly recognize premises in any order gives agents the ability to reason about existing premises in different ways, possibly leading to different theories which could then be judged along various dimensions.

Another effect of allowing multiple theories to coexist in REVOLVER is that the same premise might be revised in different ways by different users. Each user can use a unique evaluation function in the new system, embodying different revision strategies and preferences; this means that two users might not revise a premise in the same way. Hence, more than one revision might now be added to the database, although only one of these revisions would be believed by each user. While the same general hill-climbing strategy is still used for each user, each may take unique paths, leading to multiple theories.

Integrating Degrees of Belief and the Evaluation Function

Each user determines the order in which the program recognizes premises. There are two ways of accomplishing this. The user can simply recognize each premise incrementally, waiting for processing to stop before a new premise is recognized. The other way is to explicitly set the degree of belief of each premise; the system would then process premises with highest belief first. Concerning the revision process, degrees of belief can be integrated into REVOLVER's existing

[2] A later section discusses a plausible exception to this.

evaluation function as well. Just as other biases are part of this function, a bias concerning which assertions are held more strongly than others can also be used. That is, while the other measures try to implicitly decide which premise is best suited for revision, the degree of belief measure can be used to explicitly bias the system for or against the revision of certain premises.

Postulating New Objects and Models

Another new enhancement to the system involves the ability to postulate the existence of new objects (e.g., substances) in premises during the revision process. In fact, the use of unknowns can form the basis for a more general approach to revision generation: whenever additions to premises are proposed during revision, it should be possible to simply use an unknown symbol (e.g., X) and then try to resolve the identity of this unknown after further inferencing.

Resolving an unknown can sometimes be done by comparing it to other beliefs in the database. For example, if an unknown X has components C..., while a known object M also has components C..., the system can plausibly infer that X == (is equivalent to) M. Another less certain method of resolving unknowns can come by declaring two unknowns as equivalent when such an act would lead to a new model being inferred. For example, if unknown X has components C..., while unknown Y = M, the system could infer that Y == X, leading to the new model M = C....

I mentioned that theories are generally mutually independent. However, it can be useful to use beliefs inferred during the construction of one theory to influence the construction of a new theory. For example, if the components of an unknown object X in theory1 match the components of a known object M in theory2 (i.e., X = C... and M = C...), then a plausible inference would be to equate X with M, even though M = C... was never inferred during the making of theory1.

Finally, note that when the same inferences are made via different methods, the plausibility of such inferences should increase (i.e., it should become less risky to make such inferences). For example, the inference that M = C... is such a case (from the scenario just presented above).

EXAMPLE

Let us now look at a more complete example of the new ideas outlined above. We saw earlier how the original REVOLVER would handle a case from the history of chemistry involving three initial premises. However, there are order effects in this example; processing premises in a different order can lead to another theory being constructed. Thus, this example (which models the 18th century dispute between chemists Gay-Lussac and Thenard and their contemporary Davy) serves as a good example of the new concepts being presented in this paper. First we will see how two different theories can arise from the same premises, illustrating new concepts along the way. Second, we will see how these theories compare along different dimensions.

Gay-Lussac and Thenard vs. Davy

The example presented at the start of this paper essentially captures the theory held by Gay-Lussac and Thenard. The inference process in the enhanced version of REVOLVER can be represented as follows:

Meta-Assertions: Number and Explanation:
GLandT(K W → P H W) 1 (premise)
GLandT(K → P H) 2 (reduction)
GLandT(P W → K O) 3 (premise)

GLandT(K = P H) 4 from 2 (infer-components)
GLandT(P W → P H O) 5 from 4 and 3 (substitution)
GLandT(W → H O) 6 from 5 (reduction)
GLandT(W = H O) 7 from 6 (infer-components)
GLandT(P → K O) 8 (premise)
GLandT(P → P H O) 9 from 8 and 4 (substitution)
GLandT(nil → H O) 10 from 9 (reduction)

At this point, the system must perform revision. One of the revisions generated explains the inconsistent reaction by postulating an unknown substance in the inputs of 8 (i.e., P X → K O). Suppose this revision is chosen as best. Inferencing would then proceed as follows:

GLandT(P X → K O) 11 (revision)
GLandT(P X → P H O) 12 from 11 and 4 (substitution)
GLandT(X → H O) 13 from 12 (reduction)
GLandT(X = H O) 14 from 13 (infer-components)
GLandT(X == W) 15 from 14 and 7 (equate-models-with-same-components)

This last inference is made when the system tries to resolve X, and finds that water has the same components as X. Hence, X is inferred to be equivalent to W. In other words, the unknown substance in the inputs of Davy's observation P → K O is deemed to be water in the theory of Gay-Lussac and Thenard. This agrees with what took place historically (Zytkow & Simon, 1986).

However, the new system can also model the reasoning Davy used in forming his counterargument to the theory proposed by his two colleagues. The difference in reasoning begins with which premises Davy would hold with highest belief, starting with his observation P → K O. He would then process the other two premises one at a time, trying to fit each into his evolving theory:

Meta-Assertions: Number and Explanation:
Davy(P → K O) 1 (premise)
Davy(P = K O) 2 (infer-components)
Davy(K W → P H W) 3 (premise)
Davy(K → P H) 4 (reduction)
Davy(K → K O H) 5 from 4 and 2 (substitution)
Davy(nil → O H) 6 from 5 (reduction)

Now revision must be performed; the scenario is similar to the case seen earlier for Gay-Lussac and Thenard. Again, one of the revisions proposes an unknown (call it Y) in a premise's inputs to explain the inconsistent reaction. This revision is K W Y → P H W. Suppose this is selected as the best revision to make. Inferencing now proceeds:

Davy(K W Y → P H W) 7 (revision)
Davy(K Y → P H) 8 (reduction)
Davy(K Y → K O H) 9 from 8 and 2 (substitution)
Davy(Y → O H) 10 from 9 (reduction)
Davy(Y = O H) 11 from 10 (infer-components)

At this point, the next inferencing step depends on how much interaction between theories is allowed to occur (a characteristic that should depend on the agent doing the inferencing). If an agent is aware of inferences made within other theories, this can sometimes facilitate inferences within *his* theory that would not otherwise be possible. For example, in our current situation, the

system could realize that a model for water has been inferred within another theory (i.e., W = H O within the theory representing Gay-Lussac and Thenard), and notice that W's components match those of the unknown Y. Hence, Y could be equated to W, thus resolving Y.

Let us suppose that such interaction is not used, and see how inferencing proceeds. The third premise is now finally processed:

Davy(P W → K O) 12 (premise)
Davy(K O W → K O) 13 from 12 and 2 (substitution)
Davy(O W → O) 14 from 13 (reduction)
Davy(W → nil) 15 from 14 (reduction)

At this point, the system must perform revision again. One of the revisions generated explains the inconsistent reaction by postulating yet another unknown substance (call it Z), this time in the outputs of 12: P W → K O Z. Let us suppose that this revision is chosen as best; inferencing would then proceed as follows:

Davy(P W → K O Z) 16 (revision)
Davy(K O W → K O Z) 17 from 16 and 2 (substitution)
Davy(O W → O Z) 18 from 17 (reduction)
Davy(W → Z) 19 from 18 (reduction)
Davy(Z = W) 20 from 19 (infer-components)
Davy(Z == Y) 21 (assume-equivalent-unknowns)
Davy(W = H O) 22 from 21, 20 and 11 (equate-unknowns'-components)

These last two inferences show the other way in which unknowns can be resolved: by trying to equate unknowns to each other. Such an assumption is useful if new models can be inferred as a result, and this is indeed the case here. That is, when the unknowns Y and Z are equated in this example, the system can then equate the unknowns' respective components; since (11) Y = O H, and (20) Z = W, the system infers from the assumption (21) Z == Y that (22) W = H O. Note that this is the same model for water that would have resulted earlier, if the system had equated Y with the W in Gay-Lussac and Thenard's model for water. The fact that two forms of inferencing would lead to the same result makes the belief in W = H O even more plausible here (in the theory representing Davy).

Note that three different types of beliefs are all existing simultaneously in the database: the assertions, the meta-assertions representing beliefs of Gay-Lussac and Thenard, and the meta-assertions representing beliefs of Davy.

Comparing Theories

Once multiple theories can coexist in the system's database, it becomes desirable to develop criteria for comparing such theories. In the example just presented, we can make some intuitive judgments. The theory representing Gay-Lussac and Thenard seems more plausible on at least three dimensions. First, only one revision was needed to reach consistency, while Davy's theory needed two. Second, the model for water was directly inferred in the former theory, while it had to be inferred indirectly in the latter theory. Third, the former theory simply required less inferencing steps overall. Also note that these disparities between the two theories become even greater when other beliefs are added (e.g., further revisions are needed when other premises are integrated into Davy's theory, whereas they can be integrated without revision into his colleagues' theory).[3]

[3]Adding two more 18th century observations (potassium and ammonia reacting to form hydrogen and a substance

Indeed, the view of Gay-Lussac and Thenard did win out historically, and it was found that Davy's observation did overlook the presence of water in his input substances.

DISCUSSION

We have seen that while the new version of REVOLVER still hill climbs for each individual agent (i.e., only one theory is ever kept for any agent), several theories can now coexist in the system's database. Assumption-based truth maintenance systems, or ATMS (de Kleer, 1984), embody a similar approach. However, neither the ATMS nor similar systems (Doyle, 1979) address the issues of generating and selecting plausible revisions (e.g., in reaction-oriented domains such as chemistry and physics). REVOLVER was originally designed to handle both of these tasks, as well as the problem solving tasks involved in scientific discovery. Another approach to multiple theories in scientific reasoning is the ECHO system (Thagard, 1988), which models how multiple theories can develop from evidence, and how the coherence of each theory can be measured in order to evaluate and compare such theories. However, as is the case with TMS systems, ECHO does not address how or when revisions to evidence should be made. Still, Thagard's techniques for judging the explanatory coherence of theories may serve as a valuable guide for future improvements in the REVOLVER framework.

Concerning the future state of the system, the integration of degrees of belief still remains to be implemented. In addition, automatic belief in newly inferred assertions was assumed in this paper; one could imagine a more cautious mode whereby new inferences could be added to the database, but an agent would have to explicitly acknowledge that he/she wished to believe them. This cautiousness could also apply to newly created revisions; in fact, in this latter case such caution might be even more appropriate. In summary, by providing a framework for discovery, revision, and now multiple theory creation, the REVOLVER system seems to provide a solid foundation for the continued exploration of how science evolves.

REFERENCES

de Kleer, J. (1984). Choices without backtracking. *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.

Rose, D. (1988). Discovery and belief revision via incremental hill climbing. *Proceedings of the International Workshop on Machine Learning, Meta-Reasoning and Logics* (pp. 129–145). Sesimbra, Portugal.

Rose, D. & Langley, P. (in press). A hill-climbing approach to machine discovery. In *Proceedings of the Fifth International Conference on Machine Learning.*

Thagard, P. (1988). The conceptual structure of the chemical revolution. Unpublished manuscript.

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

called green-solid (K A → H G), plus another observation (G W → P A W)) leads to this effect.