18 Learning by Correcting Mistakes

In this chapter, you learn how a procedure can *repair previously acquired knowledge* by exploiting its own errors. You see that this ability is necessary because to become an expert a program must handle an ever-growing list of unusual cases, and so is forced to do knowledge repair.

In particular, you learn about FIXIT, a procedure designed to resonate with the healthy attitude that errors are valuable knowledge probes, not failures. FIXIT *isolates suspicious relations* that may cause failures, *explains why* those facts cause problems, and ultimately *repairs knowledge.*

By way of illustration, you see how FIXIT deals with a model that incorrectly allows pails to be identified as cups. FIXIT notes that the reason pails are not cups must have something to do with the handles, that cups have to be oriented, that fixed handles enable orientation, and that cups therefore have to have fixed handles, rather than hinged ones.

ISOLATING SUSPICIOUS RELATIONS

In this section, you learn how to use failures to zero in on what is wrong with an identification model. You also learn how to make superficial, ad hoc corrections.

Cups and Pails Illustrate the Problem

The cup-identification model that MACBETH learned in the previous chapter can be used to identify a variety of cups, including porcelain cups and metal cups. Unfortunately, it also allows a variety of pails, including metal pails and wooden pails, to be identified as cups. Such failures lead to the following questions:

- How can a procedure use identification failures to *isolate* suspicious relations that should perhaps prevent a model from being misapplied?
- How can a procedure use precedents to *explain* why those now isolated suspicious relations should prevent a model from being misapplied?
- How can a procedure use explanations to *repair* a model, preventing further misapplication?

Near-Miss Groups Isolate Suspicious Relations

If a pail differs from a cup only in the way the handle is attached, then the pail can act as a near miss. Unfortunately, there may be many differences, both relevant and irrelevant.

Assume, for example, that the porcelain cup in figure 18.1 is known to be a cup, and the metal pail in figure 18.1 is known to be something else, either because a teacher says so or because an attempt to drink from it fails. The pail is metal, but the cup is porcelain; the metal pail is gray, but the porcelain cup is white with balloons painted on the side; perhaps the metal pail carries water, but the porcelain cup carries coffee.

Now assume that the tin cup in figure 18.1 is also known to be a cup, but the wooden pail in figure 18.1 is known to be something else. Each pail differs from the porcelain cup in many ways, and similarly, each pail differs from the tin cup in many ways. It is important, however, that there are fewer ways in which *both* the metal pail and the wooden pail differ from *both* the porcelain cup and the tin cup. In fact, in the example, you are to assume that the cups have fixed handles, whereas the pails have hinged handles, and that nothing else characterizes the way both pails differ from both cups.

Because the model allows all four objects to be identified as cups, when the model is viewed as an antecedent–consequent rule, the antecedent relations must lie in the intersection of all the relation sets describing those four objects. Similarly, the relations, if any, that distinguish the true-success situations from the false-success situations must lie in the union of the two relation subsets shown shaded in figure 18.1.

Because the relations in the true-success set and the false-success set are likely candidates for forming explanations, they are called **suspicious relations**. Also, the situations used to identify the suspicious relations constitute a **near-miss group**, because the situations work together as a group to do the job performed by a single example and a single near miss of the sort discussed in Chapter 16.

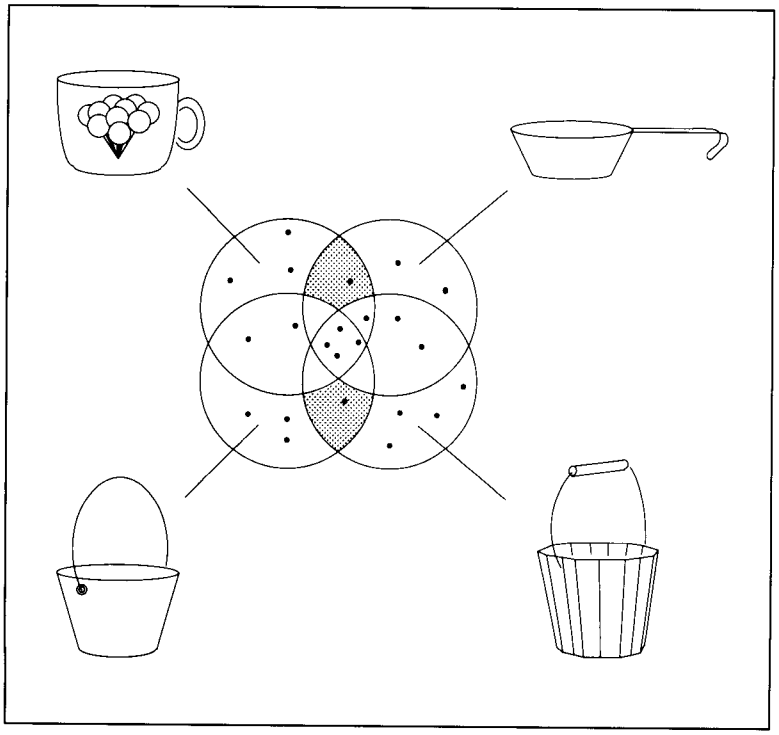

Figure 18.1 A near-miss group. The dots represent relations. Suspicious relations are in the shaded area. Some are suspicious because they are in all the true successes, but are not in any of the false successes; others are suspicious because they are in all the false successes, but are not in any of the true successes.

Clearly, isolating suspicious relations is just a simple matter of using set operations on the relations that appear in the true successes and false successes. Here is how the repair procedure, FIXIT, does the work:

To isolate suspicious relations using FIXIT,

- ▷ To isolate the true-success suspicious relations,
  - ▷ Intersect all true successes. Call the result $\bigcap T$.
  - ▷ Union all false successes. Call the result $\bigcup F$.
  - ▷ Remove all assertions in the union from the intersection. These are the true-success suspicious relations, written mathematically as $\bigcap T - \bigcup F$.
- ▷ To isolate the false-success suspicious relations,
  - ▷ Intersect all false successes. Call the result $\bigcap F$.
  - ▷ Union all true successes. Call the result $\bigcup T$.
  - ▷ Remove all assertions in the union from the intersection. These are the false-success suspicious relations, written mathematically as $\bigcap F - \bigcup T$.

In general, there will be more than one suspicious relation, but the more true successes and false successes you have, the fewer suspicious relations there are likely to be.

Suspicious Relation Types Determine Overall Repair Strategy

In the example, the *handle is fixed* relation is found in both cups, but it is *not* found in either pail. If a procedure could find a way to include this relation in the model, then the model would enable more discriminating identification. Cups would still be identified, but pails would not be.

Thus, an explanation-free repair would be to include the *handle is fixed* relation in the rule view of the model as a new antecedent condition:

Repaired Cup-Identification Rule

If	The object has a bottom
	The bottom is flat
	The object has a concavity
	The object is light-weight
	The object has a handle
	The handle is fixed
Then	The object is a cup

As described, however, the model repair is ad hoc because there is no explanation for why the new antecedent condition should work.

INTELLIGENT KNOWLEDGE REPAIR

In this section, you learn how FIXIT can explain why a particular relation makes an identification rule go wrong, thus enabling FIXIT to make informed repairs.

The Solution May Be to Explain the True-Success Suspicious Relations

Figure 18.2 shows the And tree form of the original, faulty model; figure 18.3 shows the And tree form of the model once repaired by FIXIT.

Comparing the And trees of the faulty and repaired models, you see that the *handle is fixed* relation, which is common to the true successes, now appears in the repaired model tree. There are also two other new relations: *object is manipulable* and *object is orientable.*

The old model was too general, because you cannot be sure you can drink from an object just because it carries liquids and is liftable—it has to be orientable by virtue of having a fixed handle. The new model still allows cups to be identified properly, but the increased specificity of the model prevents pails from being identified as cups.

To make the repair, FIXIT does a breadth-first reexamination of all the relations in the model's And tree, looking for a relation with an explanation

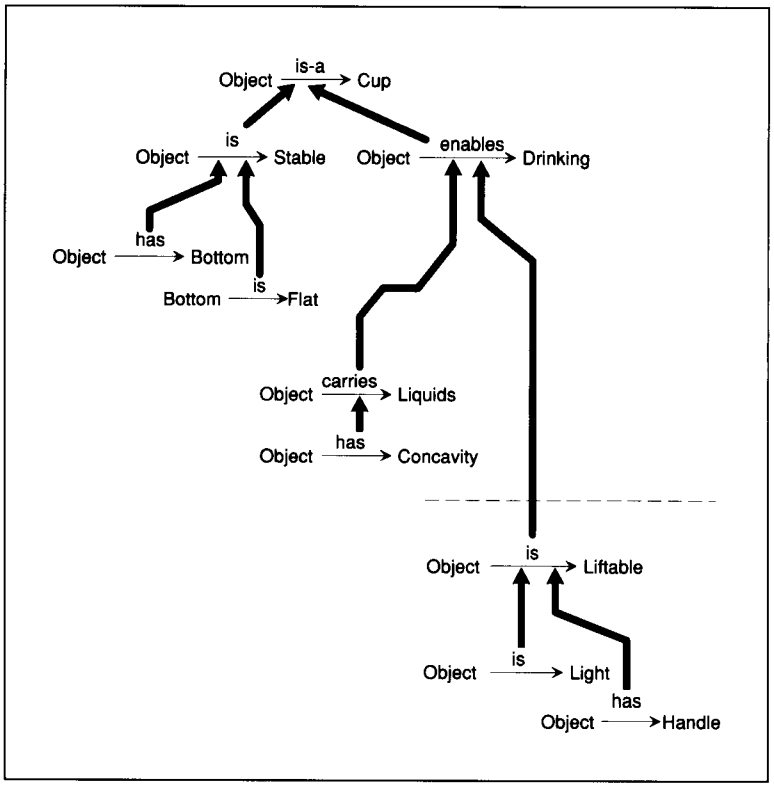

Figure 18.2 A model's And tree before repair. Both cups and pails are identified as cups. The portion of the And tree below the dashed line must be replaced.

that needs to be replaced. For each relation reexamined, FIXIT looks for precedents that tie the reexamined relation to at least one of the true-success suspicious relations. If such precedents are found, FIXIT replaces the subtree beneath the reexamined relation using those precedents, thus explaining the reexamined relation in a new way.

The new explanation should be as short as possible, because the longer the chain of precedent-supplied Cause links, the less reliable the conclusion. After all, the contributing precedents supply Cause links that are only likely; they are not certain. Consequently, FIXIT initially limits its reexamination effort to the following precedents:

- The precedents MACBETH originally used to learn the model. These are included in the expectation that much of the model will be unchanged, and therefore will be constructable from the original precedents. These original precedents constitute the initial **head set**, so called because they lie at the head end of the chain of Cause links that eventually connects the model to one or more suspicious relations.
- Those precedents in which one of the true-success suspicious relations causes something. These precedents constitute the initial **tail set**, so called to contrast with the head set.

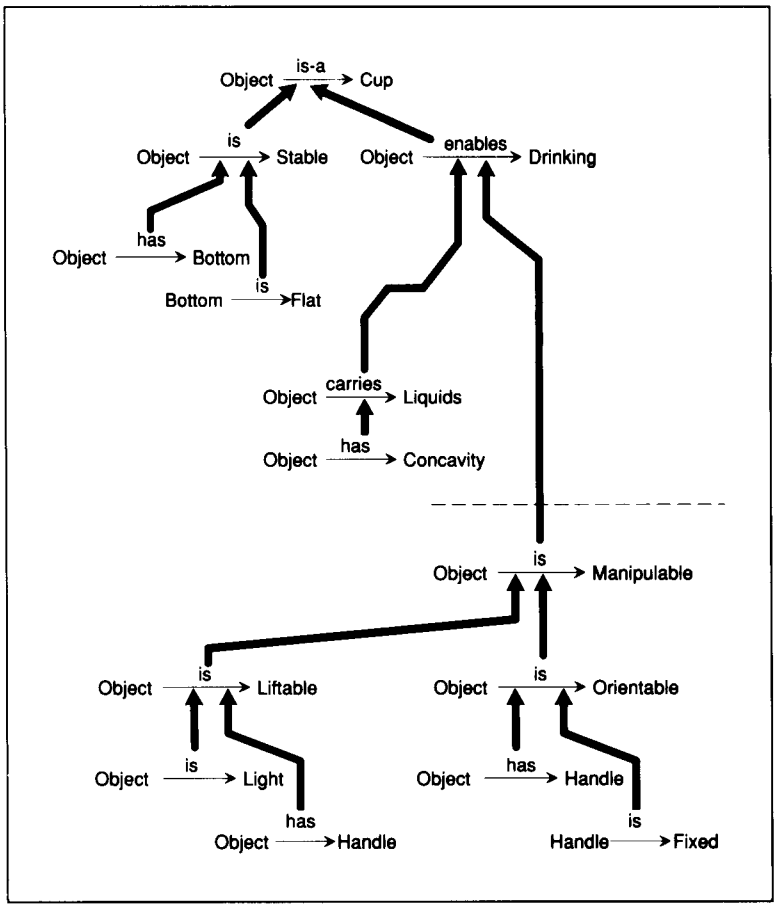

Figure 18.3 A model's And tree after repair. The repaired model allows only cups to be identified as cups. The portion of the And tree below the dashed line has been replaced.

When FIXIT reexamines a relation, it looks for a way to explain that relation using all but one of the precedents in the combined head and tail sets. The exception is the precedent that was used to explain the reexamined relation previously. That precedent is omitted so that FIXIT can explore the hypothesis that it provided an incorrect explanation, leading to the model's defective behavior.

In the cup-and-pail example, the head set consists of the cup's functional description, along with the brick, glass, bowl, and briefcase precedents used by MACBETH. The tail set consist of all those precedents in which the true-success suspicious relation, *handle is fixed*, causes something. Suppose that the *handle is fixed* relation appears in only the door precedent, in which it is tied by a Cause link to *door is orientable*. Then, the tail set consists of the door precedent alone.

Accordingly, when FIXIT reexamines the *object is-a cup* relation, it uses the brick, glass, bowl, briefcase, and door precedents. It does not use the

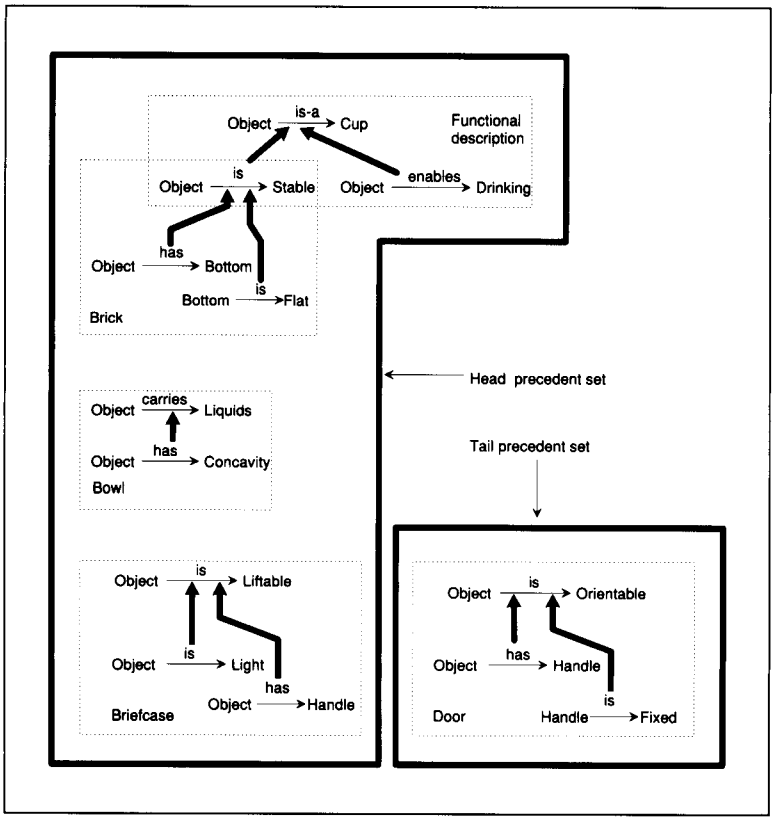

Figure 18.4 Reexamination fails because the head and tail sets are not connected to one another. Only the relevant parts of the precedents are shown.

cup's functional description because that is the precedent MACBETH used to explain *object is-a cup* when the model was learned.

At this point, FIXIT's reexamination fails to lead anywhere because the head and tail sets do not connect the reexamined relation, *object is cup*, to the suspicious relation *handle is fixed.*

Similarly, when FIXIT reexamines the *object is stable* relation, it uses the cup's functional description along with the glass, bowl, briefcase, and door precedents, omitting the brick precedent, but fails again. When it reexamines the *object enables drinking* relation, it uses the cup's functional description along with the brick, bowl, briefcase, and door precedents, omitting the glass precedent, but fails again. FIXIT also fails when it reexamines the *object carries liquids* and the *object is liftable* relations. Evidently, more precedents have to be used.

Incorporating True-Success Suspicious Relations May Require Search

Once FIXIT concludes that more precedents have to be considered, it augments the precedents in either the head or the tail sets.

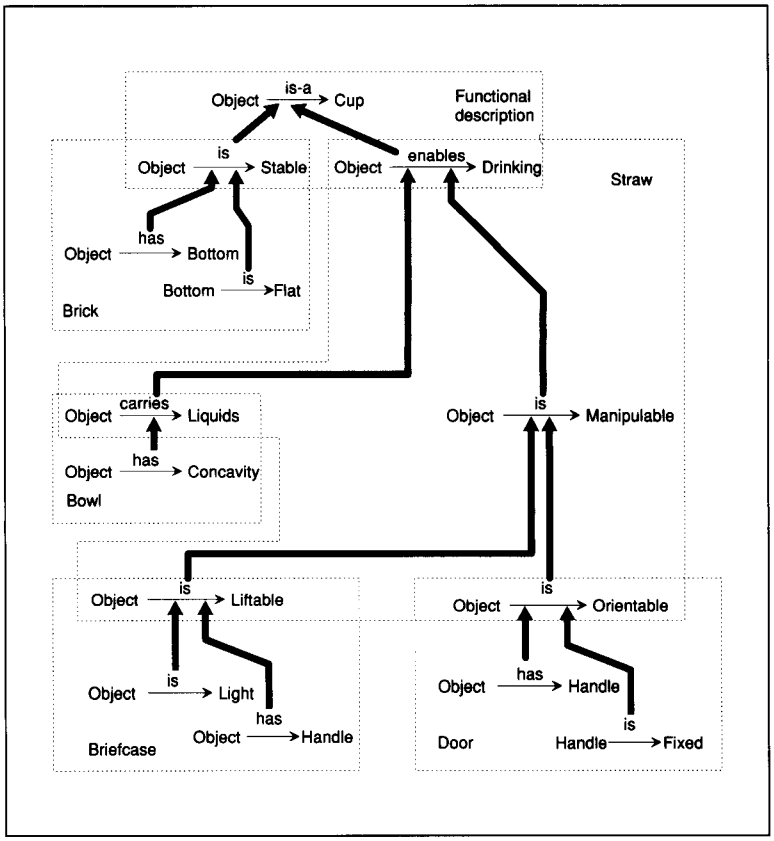

Figure 18.5 A model tree after repair with contributing precedents shown. Note that the straw precedent, having augmented the tail set, bridges the gap between the old model and the suspicious relation, *handle is fixed*. Now only cups are identified as cups; pails are not. Only the relevant parts of the precedents are shown.

To augment the head set, FIXIT looks for precedents that extend the Cause link chains that lead through the existing head-set precedents. Symmetrically, to augment the tail set, FIXIT looks for precedents that extend the Cause link chains that lead through the existing tail-set precedents.

To keep the number of precedents as small as possible, FIXIT augments only the head or the tail set, whichever has fewer precedents. In the example, FIXIT augments the tail set because it currently has only one precedent, the door precedent. FIXIT adds the straw precedent because *handle is fixed* causes *door is orientable* in the existing door precedent and *straw is orientable* causes *straw is manipulable* in the new straw precedent.

Now FIXIT starts over using the augmented tail set. As before, reexamination fails on the *object is-a cup* relation, and on the first of the relations in the next layer, *object is stable*. But when FIXIT reexamines the *object enables drinking* relation, it succeeds, connecting *object enables drinking* with the *handle is fixed* relation via the new straw precedent and the existing door precedent, as shown in figure 18.5. Of course, all the precedents shown in the figure contain details that are not shown so as to avoid clutter.

Note that the model's And tree is restructured as necessary without reasoning explicitly about wrong or missing nodes and branches. The restructured And tree can be viewed as the following antecedent–consequent rule:

Repaired Cup-Identification Rule

If	The object has a bottom
	The bottom is flat
	The object has a concavity
	The object is light-weight
	The object has a handle
	The handle is fixed
Then	The object is a cup

FIXIT stops as soon as the repaired rule no longer matches the false successes, as the following indicates:

To deal with true-success suspicious relations using FIXIT,

- ▷ Until the false successes are accounted for,
  - ▷ For each relation in the model tree, starting with the root relation and working down, breadth first,
    - ▷ Form a head set consisting of all precedents that contributed to the malfunctioning model tree.
    - ▷ Form a tail set consisting of all precedents in which one or more true-success suspicious relations cause another relation.
      - ▷ Try to find a new explanation for a relation in the model tree using the head set, minus the precedent that explained the relation before, plus the tail set.
      - ▷ If an explanation is found, replace that portion of the model that lies beneath the newly reexplained relation.
    - ▷ Augment the smaller of the head set and the tail set with other precedents that extend the causal chains found in the set's existing precedents.

The Solution May Be to Explain the False-Success Suspicious Relations, Creating a Censor

The repaired model tree works on all the cup-versus-pail problems, because not one of the pail descriptions contains a *handle is fixed* relation. There remains a danger, however, that a precedent-oriented identification procedure, such as MACBETH would try hard to explain why a pail has a fixed

handle, even though it already knows that the pail's description contains a *handle is hinged* relation.

Fortunately, FIXIT can also build recollections such as the following one, expressed as an antecedent–consequent rule:

Hinged-Handle Censor
If Handle is hinged
Then Handle is not fixed

Once this recollection is created, MACBETH, or any other identification procedure, can use the recollection to stop itself from trying to explain a relation that can be shown to be false by a single Cause link. As before, cups are identified as cups, and pails are not, but now the new recollection blocks useless effort directed at explaining that hinged-handle pails have fixed handles. A recollection so used is called a **censor**.

FIXIT creates censors if it cannot account for the false successes using the true-success suspicious relations. To create censors, FIXIT does a breadth-first reexamination of all the relations in the repaired model tree, looking for precedents that tie the negation of each reexamined relation to a false-success relation. The resulting explanation establishes why the false-success suspicious relation should block identification. This explanation, in turn, permits the creation of a new censor.

Initially, the precedent set is limited to the following, to keep the explanation as short as possible:

- Precedents in which the negation of the reexamined relation is caused by something. These precedents constitute the initial *head set.*
- Precedents in which one of the false-success suspicious relations causes something. These precedents constitute the initial *tail set.*

Here, the idea is to find an explanation for the negation of the reexamined relation that includes at least one of the false-success suspicious relations. If FIXIT finds such a collection of precedents, it creates a new censor from that collection of precedents.

Eventually, FIXIT's breadth-first reexamination tries to explain the *handle is not fixed* relation, given the false-success suspicious relation, *handle is hinged.* At this point, FIXIT uses the suitcase precedent, which happens to find its way into both the head and tail sets, because *handle is hinged* is connected to *handle is not fixed* by a cause relation:

A Suitcase

This is a description of a suitcase. The suitcase is liftable because it is has a handle and because it is light. The handle is not fixed because it is hinged. The suitcase is useful because it is a portable container for clothes.

With the suitcase precedent, FIXIT has what it needs to generate the appropriate censor.

As before, FIXIT stops as soon as the repaired rule no longer matches the false successes, as the following indicates:

To deal with false-success suspicious relations using FIXIT,

- ▷ Until the false successes are accounted for,
  - ▷ For each relation in the model tree, starting with the root relation and working down, breadth first,
    - ▷ Form a head set consisting of all precedents in which the negation of the relation is caused by another relation.
    - ▷ Form a tail set consisting of all precedents in which one or more false-success suspicious relations cause another relation.
      - ▷ Try to find an explanation for the negation of the relation in the model tree using the head set plus the tail set.
      - ▷ If an explanation is found, create a new censor.
    - ▷ Augment the smaller of the head set and the tail set with other precedents that extend the causal chains found in the set's existing precedents.

Failure Can Stimulate a Search for More Detailed Descriptions

If there are no suspicious relations in a near-miss group, there are still several ways to correct the situation:

- Although the lack of suspicious relations indicates that there is no common explanation for failure, there may be just a few explanations, each of which is common to a subset of the true successes or the false successes. The problem is to partition situations into groups, inside each of which there is a consistent explanation for failure.
- Assume that a relation that occurs in some of the true successes, but not all of them, is suspicious. See whether that relation can be used to explain the failures. If it can be, ask whether that true-success relation was left out of the other true-success descriptions by oversight.
- The lack of any suspicious relations may indicate that the situations must be described at a finer grain, adding more detail, so that an explicit explanation emerges.

Many good teachers ask their students to describe problems thoroughly before the students start to solve the problems. As a student, FIXIT certainly benefits from thorough descriptions, for thorough descriptions are more likely to contain relations that seem irrelevant at first, but that prove

suspicious when several problems are looked at from the perspective of near-miss groups.

SUMMARY

- Sometimes, a learning procedure can make a mistake, leading to a need for knowledge repair. MACBETH, for example, can learn to recognize cups, but initially confounds cups and pails.
- FIXIT uses near-miss groups, consisting of a set of positive examples and a set of negative examples, to isolate suspicious relations.
- Knowledge repair may require FIXIT either to explain true-success suspicious relations, altering an original explanation, or to explain false-success suspicious relations, creating a censor.
- A frustrated attempt to repair knowledge suggests that the right suspicious relations are not in the descriptions, and suggests that more effort should go into creating more detailed descriptions.

BACKGROUND

The near-miss group idea was conceived by Patrick H. Winston; it was first put to use by Kendra Kratkiewicz [1984] when she showed that near-miss groups could pull informative differences out of complicated descriptions of historic conflicts. Subsequently, Satyajit Rao showed how such informative differences could be used to do informed knowledge repairs [Winston and Rao 1990]. Boris Katz suggested the idea of postulating oversights when no suspicious relation emerges from a near-miss group.

The idea of using differences to focus learning appears in the work of Brian Falkenhainer [1988], in which he uses differences between working devices and nonworking devices, plus domain knowledge, to deduce why the nonworking devices fail.

The use of the term *censor* in artificial intelligence has been popularized by Marvin Minsky [1985].

One defect of the approach to learning and knowledge repair described in this chapter is that every concept has to have a name; in many experiments, however, the concepts do not correspond well to English words, forcing us to invent awkward, multiply hyphenated names. An approach to dealing with this defect is explained by Rao [1991].