ARTIFICIAL INTELLIGENCE LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MIT AI Memo 678

May 1982

Revised September 1982
Revised February 1984

LEARNING BY AUGMENTING RULES AND ACCUMULATING CENSORS

by
Patrick H. Winston

Abstract

This paper is a synthesis of several sets of ideas: ideas about learning from precedents and exercises, ideas about learning using near misses, ideas about generalizing if-then rules, and ideas about using censors to prevent procedure misapplication.

The synthesis enables two extensions to an implemented system that solves problems involving precedents and exercises and that generates if-then rules as a byproduct. These extensions are as follows:

- If-then rules are augmented by *unless* conditions, creating *augmented if-then rules.* An augmented if-then rule is blocked whenever facts in hand directly demonstrate the truth of an unless condition. When an if-then rule is used to demonstrate the truth of an unless condition, the rule is called a *censor*. Like ordinary augmented if-then rules, censors can be learned.
- Definition rules are introduced that facilitate graceful refinement. The definition rules are also augmented if-then rules. They work by virtue of *unless* entries that capture certain nuances of meaning different from those expressible by necessary conditions. Like ordinary augmented if-then rules, definition rules can be learned.

The strength of the ideas is illustrated by way of representative experiments. All of these experiments have been performed with an implemented system.

This research was done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's artificial-intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505.

KEY IDEAS

This paper builds primarily on a previous paper that introduced a theory of learning from precedents and exercises using *constraint transfer* [Winston 1981]. The theory addresses the analogy process at work when we exploit past experience in fields like Management, Political Science, Economics, Medicine, and Law, as well as from everyday life.

Two extensions to the theory are described. Work on the first extension was stimulated by some of the apparent blunders of the extant system. Work on the second extension was stimulated by some problems encountered in making definitions.

After a brief review of the overall theory, I present an example showing that the rules generated by the unextended learning system can be misapplied. Next, I discuss various solutions to the misapplication problem, including the introduction of censors. At this point *augmented if-then rules* are discussed. Each augmented if-then rule contains not only *if* and *then* parts, but also an *unless* part. Before a rule acts, censors determine if any existing facts directly demonstrate that an *unless* relation is true. If so, the rule is *blocked.*

This leads to the development of definition rules based on augmented if-then rules and a discussion of their relevance to the problem of concise definition versus unlimited nuance.

Next, it is shown that censors can block censors and that censors can be learned, both by precedent and exercise and by near miss.

Finally, I describe precedents for this work itself, including ideas that stimulated what I have done, such as Minsky's ideas on the role of censors in problem solving [Minsky 1980], as well as other ideas that I reinvented or borrowed from as my work progressed, such as Goldstein and Grimson's ideas on generalizing if-then rules [1977].

There are references throughout to an implemented system that actually does acquire and use censors. This implemented system inherits some key ingredients from previous work:

- *Analogy-based reasoning using constraint transfer.* Analogy requires the ability to determine how two situations that are similar in some respects may be similar in other respects as well. Here the determination is done by transferring constraints from the precedent situation to the exercise situation.
- *Learned if-then rules.* In contrast to current practice in Knowledge Engineering, if-then rules emerge automatically as problems are solved. Teachers supply precedents and exercises, leaving the work of formulating the if-then rules to the system.
- *Rule-based reasoning.* Once learned, rules can be used. Since rules are viewed as simple situations, the constraint-transfer programs that work with the precedent situations also work with rules, doing simple rule-based reasoning. Thus learning and reasoning reside together harmoneously in the same system.

- *Actor-object representation.* Situations are represented using relations. Each relation has true, false, or unknown as its truth value, and any relation can be an object involved in another relation.
- *Importance-dominated matching.* The similarity between two situations is measured by finding the best possible match according to what is important in the precedent situation. A precedent relation is considered important if it is connected to another relation by an importance-determining constraint. At the moment, causal connection is the only importance-determining constraint recognized.

WHAT IS TO BE UNDERSTOOD

Let us begin by reviewing the sort of task performed by the theory as previously reported. Consider the following precis of *Macbeth*, given by a teacher as a precedent:

> MA is a story about Macbeth, Lady-macbeth, Duncan, and Macduff. Macbeth is an evil noble, Lady-macbeth is a greedy, ambitious woman, Duncan is a king, and Macduff is a noble.
>
> Lady-macbeth persuades Macbeth to want to be king because she is greedy. She is able to influence him because he is married to her and because he is weak. Macbeth murders Duncan with a knife. Macbeth murders Duncan because Macbeth wants to be king and because Macbeth is evil. Lady-macbeth kills herself. Macduff is angry. Macduff kills Macbeth because Macbeth murdered Duncan and because Macduff is loyal to Duncan.

Next, consider the following exercise:

> Let E be an exercise about a weak noble and a greedy lady. The lady is married to the noble. In E show that the noble may want to be king.

Told by a teacher that *Macbeth* is to be considered a precedent, it is announced that the precedent suggests that the noble may want to be king. Then a principle-capturing if-then rule is created suggesting that the weakness of a noble and the greed of his wife can cause the noble to want to be king. The rule looks like this, printed as an if-then rule:

```
Rule
    RULE-1
if
    [LADY-4 IS GREEDY]
    [NOBLE-4 IS WEAK]
    [[NOBLE-4 IS MARRIED] TO LADY-4]
then
    [NOBLE-4 WANT [NOBLE-4 A-KIND-OF KING]]
case
    MA
```

Internally, the rule actually contains more information because the internal representation preserves the constraint structure that the rule summarizes, thereby

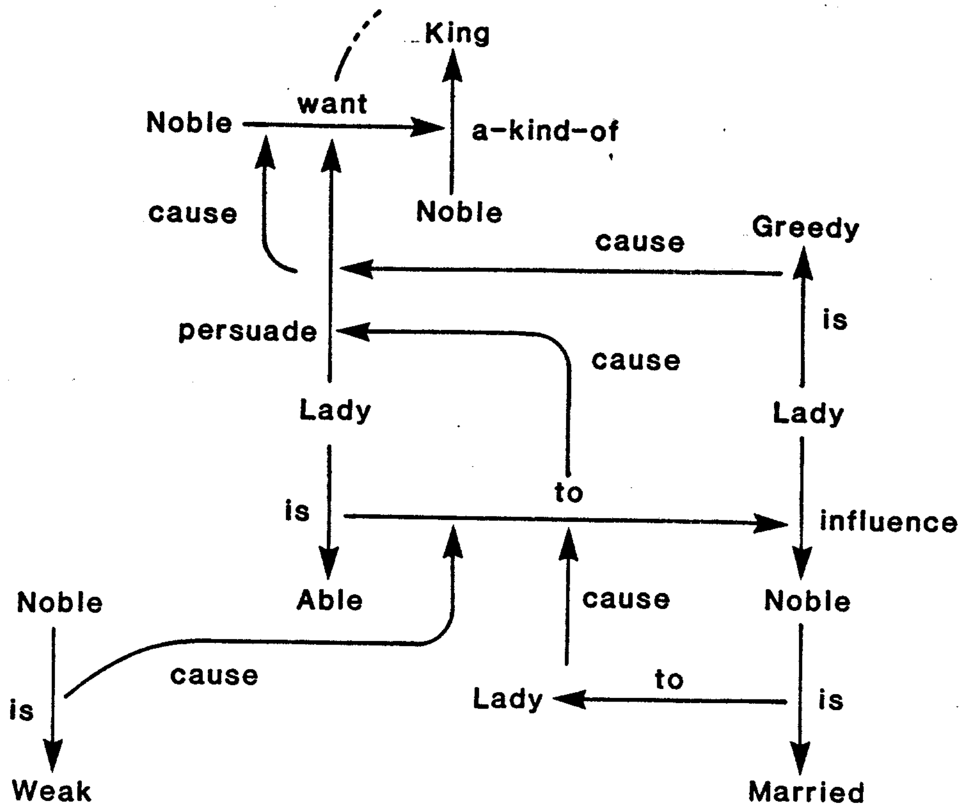

Figure 1. The internal representation of a rule. Note that the constraint structure is preserved.

recording how all the relations fit together. Figure 1 illustrates.

Here, however, it suits our purpose best to paraphrase the rules in English, as in the following third rendition of the sample rule:

Rule-1 If There is a greedy lady
and there is a week noble
and the noble is married to the lady
then the noble may want to be king

The exercise problem could have been handled by this rule directly, without recourse to the *Macbeth* precedent, were it available when the problem was posed. Thus the rule adds power. Unfortunately, it also adds blunder, as when the following exercise is given:

> Let E be an exercise about a weak noble and a greedy lady. The lady is married to the noble. He does not like her. In E show that the noble may want to be king.

This situation is different because we know that it is difficult for a person to influence someone who does not like him. Evidently, the rule is overly general, ready to reach conclusions when it should not.

This paper introduces extensions to the existing theory such that the implemented system behaves correctly on the given example and many others. To be considered a success, however, a system should not just work, it should work because it embodies arguable ideas. The arguable ideas embodied in the improved system are the following:

- The *blocking principle:* Suppose a rule, derived from a precedent, seems to apply to a problem. Consider all the relations in that part of the precedents's causal structure involved in forming the rule. If any such relation corresponds to a relation that is either false or manifestly implausible in the problem situation, then the rule based on the precedent does not apply.
- The *prima facie conjecture:* A relation is manifestly implausible if its negation can be shown by a direct, one-step inference from relations already in place.

REASONING AND CREATING RULES USING ANALOGY

Let us review how rules are generated. Consider the *Macbeth* precedent, given earlier, together with the exercise, both expressed in semantic-network form, as shown in figure 1 and figure 2.

When asked to demonstrate that the man may want to be king, given the *Macbeth* precedent, the system proceeds as follows:

- The people in the precedent are matched with the people in the exercise. More generally, precedent parts are matched with exercise parts.
- The causal structure of the precedent is mapped onto the exercise.
- It is determined that the mapped causal structure ties the relation to be shown to relations known to be true.
- A rule is constructed, with generalizations of the exercise relations used becoming *if* parts and a generalization of the relation to be shown becoming the *then* part.

When a single precedent cannot supply the total causal structure needed, the system attempts to chain several together. In the example, if it were not known already that the woman is greedy, as required for application of the *Macbeth* precedent, greed might be established through another precedent or already-learned rule. A previous paper explains this in detail [Winston 1981].

IMPROVING PERFORMANCE BY ENABLING CENSORS

So far we have established that rules can be generated and that they need to be blocked in certain circumstances. There are three obvious ways to arrange for blocking:

First, expand the *if* part of an offending rule, restricting its use. One problem with this idea is that rules can become bloated with endless tests for increasingly unlikely minutiae. Such bloat makes rules obscure and hard to criticize, debug, and improve, both for us people and for reasoning programs.

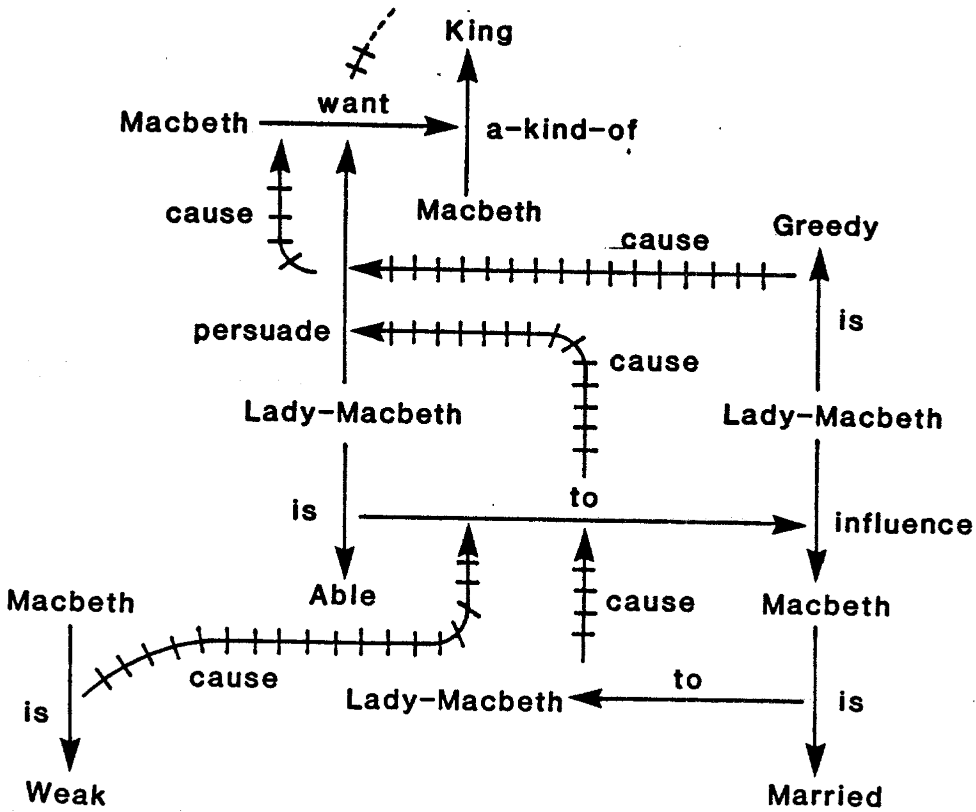

Figure 2. Problems are solved by transferring the existing cause relations of a precedent (crossed lines) onto the problem to be solved (figure 3).

Second, attach censors to each rule. Have the censors check the problem to be solved for contraindications to the rules the censors are attached to. One problem is that the rules can become bloated with censor names; these censor names would give no explicit insight into when the rules do not apply.

Third, have censors watch for particular relations. Forbid any rule or precedent to work toward establishing a relation that a censor objects to. One problem is that the rules continue to look silly, containing no hint about when they do not apply.

Censors Can Block Augmented If-then Rules

A better, less obvious idea, is this:

- Augment each rule at the time it is generated with entries that correspond to all relations in the causal structure lying between relations that enter the *if* part of the rule and the relation that enters the *then* part of the rule. Negations of these intermediate entries constitute the *unless* part of the rule. According to the blocking principle, if any entry in the *unless* part of the rule corresponds to something that is manifestly true, then the rule does not apply.

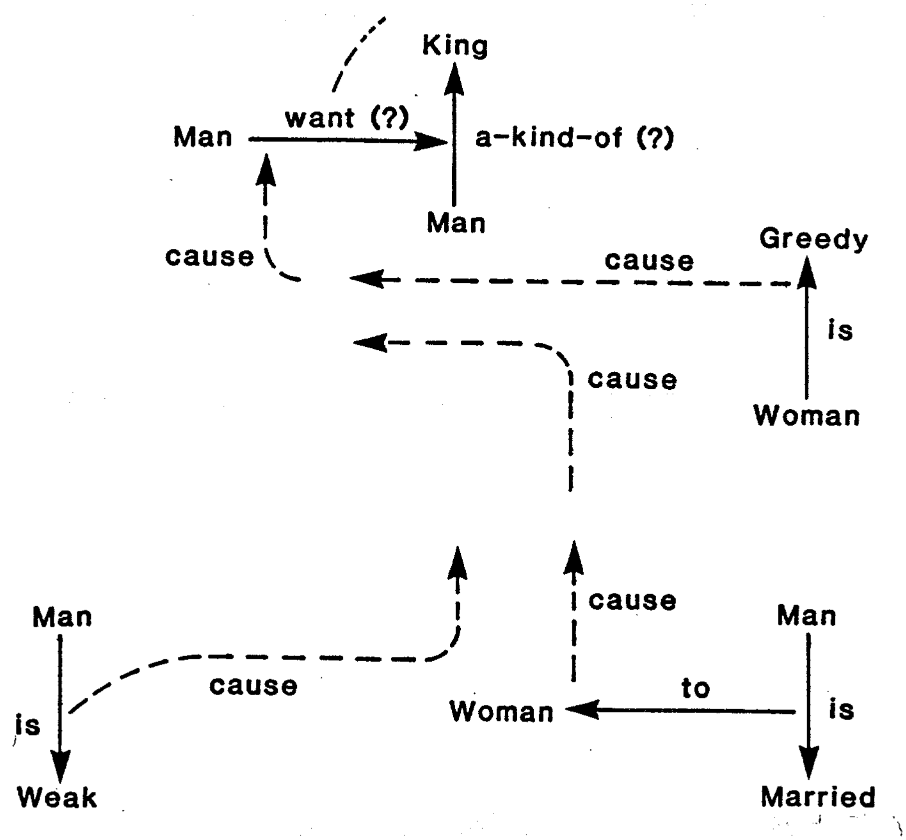

Figure 3. Problems are solved by transferring the existing cause relations of a precedent (figure 2) onto the problem to be solved (dotted lines).

Note that entries in the *unless* part of the rule are distinguished by the way that they are used, not by the fact that the truth value of many is false. In our examples, the *unless* entries usually will be false relations because they are negations of relations that are usually true. But a precedent's intermediate relations, those from which *unless* entries are made, may be false relations, leading to *unless* entries that are true relations.

Clearly a relation is manifestly true if the existing facts indicate that the relation is true. But introspectively, it seems unreasonable to go deeply into reasoning about *unless* entries.

Hence the implemented system adheres to the following specialization of the *prima facie* conjecture:

- If any entry in the *unless* part of a rule corresponds to a relation that can be shown to be true by another rule working directly from relations already in place, without further inference, then block the rule.

Note that the restriction to one-step inference is an attempt to translate *manifestly true* into a computationally precise mechanism. Doublessly there will be better translations.

Let us consider an example. Suppose, that a rule's *unless* part is triggered when someone is unable to influence another. Such a rule will be blocked if the person to be influenced does not like the other. The augmented form of RULE-1 is:

Rule-1 If There is a greedy lady
and there is a week noble
and the noble is married to the lady
then the noble may want to be king
unless the lady does not persuade the noble to want to be king
or the lady is not able to influence the noble

The blocking rule is:

Rule-2 If There is a person, x, who does not like another person, y
then y is not able to influence x

A rule becomes a *censor* when it blocks the application of another rule. Since censors look just like any other rules, censors can be learned, stored, and retrieved in the same ways.

Note that when the illustrated rule is used to block another, it only works if it is known at the time of use that there is dislike. There is no attempt to demonstrate dislike when not already known.

Note that the viability of the *prima facie* conjecture depends on having a rich vocabulary of relations. It would be difficult to demonstrate anything in one step if all relations were reduced to canonical constellations of small-vocabulary primitives. This opens the question of just how rich the vocabulary should be, a question answered operationally by using freely those relations for which there are common natural-language words.

The viability of the *prima facie* conjecture also depends on having all solid facts available before backward-chaining problem solving begins. This means that all solid facts are either given facts or deduced already by forward chaining from given facts using reliable, potentially relevant rules. Reliability is insured by forward chaining only with rules that reach unassailable conclusions. Relevance cannot be insured, but can be rendered more likely. One way is to use the context mechanism described in an earlier paper [Winston 1981].

Censors Can Block Censors

Actually, it is possible to be influenced by someone you dislike if for some reason you trust them in spite of the dislike. Perhaps the real able-to-influence censor should look like this:

Rule-2 If There is a person, x, who does not like another person, y
then y is not able to influence x
unless x trusts y

Such a censor could be blocked by another censor which states that you believe someone if they have the ability to convince you:

Censor-1 If There is a person, x, who is able to convince another person, y
then y trusts x

To illustrate how these can interact, consider the following situation:

> Let E be an exercise about a weak noble and a greedy lady. The lady is married to the noble. He does not like her. However, the lady is able to convince the noble. In E show that the noble may want to be king.

This produces the following scenario:

- First, the problem is posed and RULE-1 is fetched. Its *if* parts are satisfied.
- Next, the *unless* part of RULE-1 is examined. The line involving ability to influence causes RULE-2 to be fetched. Its *if* parts are satisfied. RULE-1 is about to be blocked.
- But RULE-2's *unless* part must be examined. The line involving believing causes CENSOR-1 to be fetched. Its *if* parts are satisfied. Thus CENSOR-1 blocks RULE-2, preventing RULE-2 from blocking RULE-1.
- Finally, RULE-1 succeeds, establishing the relation originally asked about.

Augmented If-then Rules Are Not Rules of Inference

It is tempting to write censors in the following way:

$$A_1 \,\&\, \ldots \,\&\, A_n \,\&\, \neg(B_1 \vee \ldots \vee B_n) \rightarrow C$$

or alternatively,

$$A_1 \,\&\, \ldots \,\&\, A_n \,\&\, \neg B_1 \,\&\, \ldots \,\&\, \neg B_n \rightarrow C$$

where the As are in the *if* part of the rule and the Bs are in the *unless* part.

Logical notation is deceptive, however, for in the use of augmented if-then rules, the As and Bs get treated differently from each other, in contrast to the conventions of traditional logic: unlimited effort is to be put into showing the As are true; only one-step effort is put into showing that the Bs are true, with the Bs assumed false on failure.

Note that rules used as censors are not permitted to create new objects. This insures that the amount of computation added by the application of censors to *unless* entries is bounded even though censors have their own *unless* parts that must be checked by censors. I believe it is likely that censor computations will prove in practice to be broad and shallow, as well as bounded, suggesting parallel implementation.

Augmented Rules Suggest an Approach to Certain Definition Problems

Winograd has discussed the difficulty of definition using the word *bachelor* [Winograd 1976]. To be sure, a bachelor is an unmarried adult man, but Winograd notes that such a definition can cause trouble if used when someone says, "Please invite some nice bachelors to my party," for it would be strange to invite certain kinds of bachelors. For example, Catholic priests and misogamists, while satisfying the dictionary definition, are clearly not what a party giver has in mind.

Since the exception possibilities seem limitless, Winograd feels it is inappropriate to rest a definition of *bachelor* on a clearly defined, small set of primitive propositions, arguing that it is better to think of using some abstract measure of closeness to an extensible set of exemplars. Woods takes issue with Winograd's view, feeling that correct understanding must involve an explicit selection of a particular word sense, rather than closeness to a generally applicable exemplar set [Woods 1981].

The augmented-rule idea may offer a slightly different approach to the problem. Consider the following definition of *bachelor*, stated as an augmented if-then rule:

```
Rule-2  If      There is a man
                and the man is not married
                and the man is an adult
        then    the man is a bachelor
        unless  the man is not expected to be married
                or the man is not able to be married
```

With this definition, the conclusion can be avoided, even though the *if* part of the rule is fully satisfied, providing that the individual involved is not able to be married or is not expected to be married. This takes care of the priest and the misogamist problems, given the following censors:

```
Censor-1  If    A man is a misogamist
          then  the man is not expected to be married

Censor-2  If    A man is a priest
          then  the man is not able to be married
```

Evidently, it is possible to have a simple, stable definition of bachelor, while at the same time allowing for knowledge relevant to bachelors to interact with the definition, when appropriate, as that knowledge is accumulated. As more is learned, the definition is used more intelligently, and, in a sense, the definition is never closed.

How does capturing the meaning of *bachelor* with an augmented if-then rule compare with other approaches? One point of view is that Winograd's exemplars correspond to rule-generating precedents, and learned augmented if-then rules correspond to Woods's selectable word senses. We will turn to learning about bachelors from precedents in a moment.

Censors Can Improve Precedent Reasoning

While censors were originally investigated in this work in order to cure the apparent silliness of some learned rules, they help in another context too. When ordinary precedent-exercise problem solving is in progress, the analogy part of the system works back through the causal structure in the precedent, looking for relations that correspond to relations in the exercise. Each time there is no corresponding relation, before the system moves further through the causal structure, it does a censor check.

Work with a precedent stops if a censor check exposes a relation in the precedent that corresponds to a relation in the exercise that is manifestly improbable. The precedent's intended conclusion is judged inoperable because the exercise supports the censor's blocking conclusion.

LEARNING AUGMENTED RULES

Since censor rules and definition rules are just rules used in a special way, they can be learned just like any other rules. This may be by direct telling, or it may be by precedent and exercise, or it may be by near miss.

Augmented Rules Can Be Learned by Precedent and Exercise

Here is a precedent and an exercise for learning thé bachelor definition rule:

> Let S be a story about Casanova. Casanova is a bachelor because he is a man and because he is expected to be married. He is expected to be married because he is able to be married. He is able to be married because he is an adult and because he is not married.
>
> Let E be an exercise about Henry. He is a man and an adult. He is not married. In E show that Henry is a bachelor.

Of course, one might argue that providing the precedent involving Casanova is unrealistic spoon feeding. Indeed, it may well be, so it is important to understand that the same bachelor rule can be learned using several independent precedents:

> Let S be a story about a man. He is a bachelor because he is expected to be married. He is a bachelor because he is a man.
>
> Let S be a story about a man. He is expected to be married because he is able to be married.
>
> Let S be a story about a man. He is able to be married because he is an adult and because he is not married.

Alternatively, the bachelor rule can be learned using several previously-learned rules:

```
Story-1  If    There is a man
               and the man is an adult
         then  the man is expected to be married
```

Story-2	If	A man is able to be married
	then	the man is expected to be married
Story-3	If	A man is an adult
		and the man is not married
	then	the man is able to be married

Also, it is possible to learn a rule that allows a married Moslem, seeking an additional wife, to be considered a bachelor.

Augmented Rules Can Be Learned by Near-miss

Of course, there should be some way of recovering if an impoverished definition is acquired early on. The near-miss idea seems useful in such situations. Consider this scenario:

- A teacher tells the system that a bachelor is an unmarried, adult man. This produces an impoverished definition of bachelor, one without anything in the *unless* part.
- The teacher complains when the system identifies a Catholic priest as a bachelor.
- The system notices that the only robust difference between the priest and other people who are correctly identified as bachelors is that the priest is not able to be married.
- The system guesses that bachelors must be able to be married and puts an appropriate entry in the *unless* part of the bachelor definition.

Of course, this is a particularly simple situation since there is but one object involved and the descriptions are such that the near-miss-causing relation is the only relation that is caused by something and not deemed plausible in a situation where the rule does apply. It is not known how difficult it would be to identify the right difference in general, but recent work on *near-miss groups* suggests that the right difference can be identified, given that there are several situations for which the rule works and should as well as several for which the rule works but should not [winston 1984].

In the event that there is no way to narrow down the possibilities conclusively, there are two approaches, both of which are under study and deserve attention. One approach is to do search, perhaps massive search. The other alternative is to do nothing. Work by Berwick on syntax acquisition [1982] and by Minsky in concept learning [unpublished draft] both suggest that if it is difficult to identify the right difference, a learning system should simply give up, waiting for more transparent examples to come along.

THE IMPLEMENTED SYSTEM

The example precedents, exercises, rules, and censors in this paper are shown in the exact English form used by the implemented system. Translation from English into the semantic net representation used by the system is done by a parser developed and implemented by Boris Katz [Katz 1980, Katz and Winston 1982]. The grammar used by the parser is also used by a generator, which produces English versions of the rules.

So far, the system knows a few dozen censors, most of which it is told, all of which it can learn from precedents or rules and exercises. Clearly the number is enough to do surface-scratching experiments and to illustrate the ideas, but an order of magnitude or two more will be required to demonstrate the ideas.

OPEN QUESTIONS

It is plain that this work is only a beginning. Work is in progress on several related fronts:

- Exploiting several situations for which a rule works and should, together with several for which the same rule works but should not, in order to improve rule the rule.
- In collaboration with ryszard michalski: generalizing the notion of manifestly improbable in order to devise a variable-precision logic.
- In collaboration with Tomas O. Binford (Stanford University), Michael Lowry (Stanford University), and Boris Katz: creating appearance descriptions from functional descriptions, precedents, and examples [Winston et. al 1983].
- In collaboration with peter andrea: using abstractions in matching and in indexing and retrieving.
- In collaboration with richard doyle: the problem of incorporating time into the representation.
- In collaboration with Boris Katz AND OTHERS: retrieving precedents from a data base so that they need not be given by a teacher.
- In response to a suggestion by J. Míchael Brady: augmentING the rules with an *if-relevant* part in addition to the *unless* part described in this paper. The idea is that the *if-relevant* part will somehow keep track of the ultimate goals a rule may be relevant to, so that the rule is used in forward chaining only if one of the potential ultimate goals is involved in the problem to be solved. This would make the rules look like this in logical notation:

$$A_1 \mathbin{\&} \ldots \mathbin{\&} A_n \mathbin{\&} \neg(B_1 \vee \ldots \vee B_n) \mathbin{\&} (G_1 \vee \ldots \vee G_n) \rightarrow C$$

  where the *A*s are in the *if* part of the rule, the *B*s are in the *unless* part, and the *G*s are in the *if-relevant* part; and where it is understood that only one-step effort is to be put into the *B*s and *G*s. This would complement the existing context mechanism explained previously [Winston 1981].

In addition, the following open questions, enumerated in a previous paper, remain open [Winston 1981]:

- There is no way to handle subcategories of cause such as those sketched by Rieger [1978].
- There is no way to handle constraints about quantities such as those constraints that appear in the work of Forbus [1982].
- There is no way to summarize an episode in a story so as to make a general precis leading to more abstract rules. Lehnert's summarization work should be tried [Lehnert 1981].

CONCLUSION: SIMPLE IDEAS HAVE PROMISE

This paper is about a set of ideas that enable improvement in the reliability of learned rules. The extended theory enables improved performance in those domains subject to problem solving by analogy. Such domains satisfy certain restrictions:

- The situations in the domain can be represented by the relations between the parts together with the classes and properties of those parts.
- The importance of a part of a description is determined by the constraints it participates in.
- Constraints that once determine something will tend to do so again.

Things that involve spatial, visual, and aural reasoning do not seem to satisfy all the restrictions. Things that involve Management, Political Science, Economics, Law, Medicine, and ordinary common sense do seem to satisfy the restrictions, however, and are targets for the learning and reasoning ideas of the theory:

- Actor-object representation.
- Importance-dominated matching.
- Analogy-based reasoning using constraint transfer.
- If-then rules learned by solving problems.
- If-then rules improved by modifications based on near misses.
- If-then rules augmented by unless parts.
- Blocking censors that create fences around rules using *prima facie* evidence.

RELATED WORK

This work builds on the MACBETH system [Winston 1979, 1981, 1984], which concentrated on analogy and rule acquisition. Also, Minsky's views on censors had a major influence [Minsky 1980].

To a lesser extent, the idea of learning by near miss is involved [winston 1970]. However, in this newer work, there is not only a different purpose, there is a much greater degree of participation by the learning system in the learning process because many precedents, rules, and censors have to be accepted or retrieved and made to work together, not just a single model and near miss. Consequently, on the splectrum ranging from learning by being told to learning by discovery, this newer work lies farther toward the learning-by-discovery end.

The augmented if-then rule is a special case of the annotated if-then rule introduced by Goldstein and Grimson in a paper on flight simulation [1977]. They had the idea that if-then rules should exhibit certain unless-like conditions (which they called caveats), as well as rationales, plans, and control information. The work of Brown and VanLehn on explaining subtraction bugs is a more recent precedent for using censors to block rules, although their censors (which they call critics) are triggered by what a rule does, rather than by unless conditions [Brown and VanLehn 1980].

The idea that censors should work only with the facts in hand is a variant on the theme of reasoning using limited resources, an idea that is discussed widely, particularly in the expert-systems literature.

John Mallery observed in conversation that the definition of *bachelor* really should say something about being expected to be married, stimulating me to try handling the bachelor problem within the *unless* framework. Boris Katz pointed out that the *prima facie* conjecture does not make sense unless all reliable, potentially relevant forward chaining is done first. Jonathan H. Connell suggested using the designator *unless*, rather than *if-plausible*, which was used in a previous version of this paper.

And of course, analogy was first studied in artificial intelligence by Thomas G. Evans [1963], the matcher in his geometric analogy program ranked matches according to built-in, implicit importance criteria. Matches viewing two figures as rotations of one another, for example, were regarded as stronger than matches viewing the same two figures as reflections. Evans was able to use fixed, built-in importance criteria since he worked only with the world of simple figures.

ACKNOWLEDGMENTS

This paper was improved by comments from Robert Berwick, J. Michael Brady, Boris Katz, Michael Lowry, and Karen A. Prendergast.

REFERENCES

Berwick, Robert "Locality Principles and the Acquisition of Syntactic Knowledge," PhD Thesis, Department of Electrical Engineering & Computer Science, MIT, 1982.

Brown, John Seely and Kurt VanLehn, "Repair Theory: A Generative Theory of Bugs in Procedural Skills," *Cognitive Science*, vol. 4, no. 4, October-December, 1980.

Davis, Randall "Applications of Meta Level Knowledge to the Construction, Maintenance, and Use of Large Knowledge Bases," PhD Thesis, Stanford University, Stanford, California, 1979. Also in *Knowledge-Based Systems in Artificial Intelligence*, Randall Davis and Douglas Lenat, 1980.

Evans, Thomas G., "A Heuristic Program to Solve Geometric Analogy Problems," in *Semantic Information Processing*, edited by Marvin Minsky, MIT Press, 1968. Based on PhD thesis, Department of Electrical Engineering, MIT, 1963.

Forbus, Ken, "Qualitative Reasoning about Physical Processes," submitted for publication, 1982.

Goldstein, Ira P. and Eric Grimson, "Annotated Production Systems A Model for Skill Acquisition," *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, 1977.

Katz, Boris, "A Three-step Procedure for Language Generation," Artificial Intelligence Laboratory Memo No. 599, December, 1980.

Katz, Boris and Patrick H. Winston, "Parsing and Generating English using Commutative Transformations," Artificial Intelligence Laboratory Memo No. 677, May, 1982. Also see "A Two-way Natural Language Interface," in *Integrated Interactive Computing Systems*, edited by P. Degano and E. Sandewall, North-Holland, Amsterdam, 1982.

Lehnert, Wendy, "Plot Units and Narrative Summarization," *Cognitive Science*, vol 4, 1981.

Minsky, Marvin, "Jokes and the Logic of the Cognitive Unconscious," MIT Artificial Intelligence Laboratory Memo No. 603, November 1980.

Minsky, Marvin, unpublished draft on Minsky's society of mind theory of thinking.

Rieger, Chuck, "On Organization of Knowledge for Problem Solving and Language Comprehension," *Artificial Intelligence*, vol. 7, no. 2, 1978.

Winograd, Terry, "Towards a Procedural Understanding of Semantics," Stanford Artificial Intelligence Laboratory Memo AIM-292, Stanford Computer Science Department Report No. STAN-CS-76-580, November, 1976.

Winston, Patrick Henry, "Learning Structural Descriptions from Examples," PhD thesis, MIT, 1970. A shortened version is in *The Psychology of Computer Vision*, edited by Patrick Henry Winston, McGraw-Hill Book Company, New York, 1975.

Winston, Patrick Henry, "Learning by Creating and Justifying Transfer Frames," *Artificial Intelligence*, vol. 10, no. 2, 1978.

Winston, Patrick Henry, "Learning and Reasoning by Analogy," *CACM*, vol. 23, no. 12, December, 1980. A version with details is available as "Learning and Reasoning by Analogy: the Details," MIT Artificial Intelligence Laboratory Memo No. 520, April 1979.

Winston, Patrick Henry, "Learning New Principles from Precedents and Exercises," *Artificial Intelligence*, vol. 23, no. 12, 1982. A version with details is available as "Learning New Principles from Precedents and Exercises: the Details," MIT Artificial Intelligence Laboratory Memo No. 632, May 1981.

Winston, Patrick Henry, "Learning by Augmenting Rules and Accumulating Censors: the Details," MIT Artificial Intelligence Laboratory Memo No. 678, May 1982.

Winston, Patrick Henry, *Artificial Intelligence, Second Edition*, Addison-Wesley, 1984.

Winston, Patrick Henry, "Improving Learned Constraints using Symbolic Correlation," in preparation.

Winston, Patrick Henry, Thomas O. Binford, Boris Katz, and Michael R. Lowry, "Learning Physical Descriptions from Functional Definitions, Examples, and Precedents," National Converence on Artificial Intelligence, Washington, D. C., 1983.

Woods, William A., "Procedural Semantics as a Theory of Meaning," Bolt Beranek and Newman Report No. 4627, March 1981. Also available in *Computational*

Aspects of Linguistic Structures, A. Joshi, I. Sag, and B. Webber (eds.), Cambridge University Press.

APPENDIX: THE EXPERIMENTS

This appendix contains traces of the set of experiments which are described in the body of this paper.

Experiment 1:

This exercise illustrates the use of one censor to block another, preventing the blocking of a rule.

> Let E be an exercise about a weak noble and a greedy lady. The lady is married to the noble. He does not like her. The lady is able to convince the noble. In E show that the noble may want to be king.

```
I am trying to show [NOBLE-3 WANT [NOBLE-3 A-KIND-OF KING]]
Supply y, n, ?, r = rules, p = precedents, or a suggestion:
> r
I find RULE-1 +
Matching EXERCISE-4 to RULE-1 producing MATCH-22
0+
The match score is 100.  ((LADY-3 LADY-4) (NOBLE-3 NOBLE-4))
I note that [LADY-3 IS GREEDY] for use with RULE-1
I note that [NOBLE-3 IS WEAK] for use with RULE-1
I note that [[NOBLE-3 IS MARRIED] TO LADY-3] for use with RULE-1
----- start of censor check -----
++
Matching EXERCISE-4 to RULE-2 producing MATCH-23
The match score is 100.  ((NOBLE-3 PERSON-8) (LADY-3 PERSON-7))
I note that [[NOBLE-3 LIKE LADY-3] IS FALSE] for use with RULE-2
----- start of censor check -----
++
Matching EXERCISE-4 to CENSOR-1 producing MATCH-24
The match score is 100.  ((LADY-3 PERSON-6) (NOBLE-3 PERSON-5))
I note that [[LADY-3 IS ABLE] TO [LADY-3 CONVINCE NOBLE-3]]
  for use with CENSOR-1
The evidence from CENSOR-1 prevents [NOBLE-3 TRUST LADY-3]
----- end of censor check -----
Rule RULE-2 is blocked.
The evidence from RULE-2 does not prevent [[LADY-3 IS ABLE] TO [
  LADY-3 INFLUENCE NOBLE-3]]
----- end of censor check -----
The evidence from RULE-1 indicates [NOBLE-3 WANT [NOBLE-3 A-KIND-OF KING]
 ]
No need to derive a new rule.
```

Experiment 2:

The task is to learn what a bachelor is.

> Let S be a story about Casanova. Casanova is a bachelor because he is

expected to be married. Casanova is a bachelor because he is a man. He is expected to be married because he is able to be married. He is able to be married because he is an adult and because he is not married.

Let E be an exercise about Henry. He is a man and an adult. He is not married.

In E show that Henry is a bachelor.

```
I am trying to show [HENRY A-KIND-OF BACHELOR]
Supply y, n, ?, r = rules, p = precedents, or a suggestion:
> story-1
Matching EXERCISE-1 to STORY-1 producing MATCH-9
0+
The match score is 100.  ((HENRY CASANOVA))
----- start of censor check -----
+
Matching EXERCISE-1 to CENSOR-1 producing MATCH-10
The match score is 100.  ((HENRY PERSON-1))
The evidence from CENSOR-1 does not prevent [[HENRY IS MARRIED]
  IS EXPECTED]
----- end of censor check -----
----- start of censor check -----
+
Matching EXERCISE-1 to CENSOR-2 producing MATCH-11
The match score is 100.  ((HENRY MAN-4))
The evidence from CENSOR-2 does not prevent [[HENRY IS ABLE] TO [
  HENRY IS MARRIED]]
----- end of censor check -----
I note that [HENRY A-KIND-OF ADULT] for use with STORY-1
I note that [[HENRY IS MARRIED] IS FALSE] for use with STORY-1
I note that [HENRY A-KIND-OF MAN] for use with STORY-1
The evidence from STORY-1 indicates [HENRY A-KIND-OF BACHELOR]
Rule RULE-2 is derived from STORY-1 and looks like this:
Rule
    RULE-2
if
    [MAN-10 A-KIND-OF MAN]
    [[MAN-10 IS MARRIED] IS FALSE]
    [MAN-10 A-KIND-OF ADULT]
then
    [MAN-10 A-KIND-OF BACHELOR]
unless
    [[[MAN-10 IS MARRIED] IS EXPECTED] IS FALSE]
    [[[MAN-10 IS ABLE] TO [MAN-10 IS MARRIED]] IS FALSE]
case
    STORY-1
Should I index it as a rule?
> y
```

Experiment 3:

The task is to learn an alternative definition of bachelor, suited to those who may have more than one wife.

> Let S be a story about Casanova. Casanova is a bachelor because he is expected to be married. Casanova is a bachelor because he is a man. He is expected to be married because he is able to be married. He is able to be married because he is an adult and because he is not married.
>
> Let E be an exercise about a man. He is an adult and a moslem. He is married. He is able to be married again because he is a moslem.
>
> In E show that the man is a bachelor.

```
I am trying to show [MAN-7 A-KIND-OF BACHELOR]
Supply y, n, ?, r = rules, p = precedents, or a suggestion:
> story-1
Matching EXERCISE-4 to STORY-1 producing MATCH-12
+
The match score is 100.  ((MAN-7 CASANOVA))
----- start of censor check -----
+
Matching EXERCISE-4 to CENSOR-1 producing MATCH-13
The match score is 100.  ((MAN-7 PERSON-1))
The evidence from CENSOR-1 does not prevent [[MAN-7 IS MARRIED] IS
  EXPECTED]
----- end of censor check -----
I note that [[MAN-7 IS ABLE] TO [MAN-7 IS MARRIED]] for use with
  STORY-1
I note that [MAN-7 A-KIND-OF MAN] for use with STORY-1
The evidence from STORY-1 indicates [MAN-7 A-KIND-OF BACHELOR]
Rule RULE-3 is derived from STORY-1 and looks like this:
Rule
    RULE-3
if
    [MAN-11 A-KIND-OF MAN]
    [[MAN-11 IS ABLE] TO [MAN-11 IS MARRIED]]
then
    [MAN-11 A-KIND-OF BACHELOR]
unless
    [[[MAN-11 IS MARRIED] IS EXPECTED] IS FALSE]
case
    STORY-1
Should I index it as a rule?
>  y :
```

Experiment 4:

The task is to improve a simple definition of bachelor using a near miss.

> Let R be a rule about a person. The person is a bachelor because he is a man, because he is an adult, and because he is not married.
>
> Let X be a story about a man. He is an adult. He is not married.
>
> Let Y be a story about a man. He is an adult. He is not married. He is not able to be married because he is a priest.
>
> Fix rule R which applies to X but not to Y.

```
The raw rule is:
Rule
    RULE-1
if
    [PERSON-2 A-KIND-OF MAN]
    [PERSON-2 A-KIND-OF ADULT]
    [[PERSON-2 IS MARRIED] IS FALSE]
then
    [PERSON-2 A-KIND-OF BACHELOR]
Matching STORY-6 to STORY-5 producing MATCH-14
The match score is 0.  NIL
Match for MAN-9 using A-KIND-OF is MAN-8
----- start of censor check -----
+
Matching STORY-5 to CENSOR-2 producing MATCH-15
The match score is 100.  ((MAN-8 MAN-4))
The evidence from CENSOR-2 does not prevent [[MAN-8 IS ABLE] TO [
  MAN-8 IS MARRIED]]
----- end of censor check -----
Is it plausible that [[[MAN-8 IS ABLE] TO [MAN-8 IS MARRIED]]
  IS FALSE]
> n
Evidently the relevant difference is that [[[MAN-9 IS ABLE] TO [
  MAN-9 IS MARRIED]] IS FALSE] in STORY-6
Matching STORY-6 to RULE-1 producing MATCH-16
+
The match score is 100.  ((MAN-9 PERSON-2))
Rule
    RULE-1
if
    [PERSON-2 A-KIND-OF MAN]
    [PERSON-2 A-KIND-OF ADULT]
    [[PERSON-2 IS MARRIED] IS FALSE]
then
    [PERSON-2 A-KIND-OF BACHELOR]
unless
    [[[PERSON-2 IS ABLE] TO [PERSON-2 IS MARRIED]] IS FALSE]
```