LEARNING PHYSICAL DESCRIPTIONS FROM FUNCTIONAL DEFINITIONS, EXAMPLES, AND PRECEDENTS

by
Patrick H. Winston†
Thomas O. Binford‡
Boris Katz†
and Michael Lowry‡

Abstract

It is too hard to tell vision systems what things look like. It is easier to talk about purpose and what things are for. Consequently, we want vision systems to use functional descriptions to identify things, when necessary, and we want them to learn physical descriptions for themselves, when possible.

This paper describes a theory that explains how to make such systems work. The theory is a synthesis of two sets of ideas: ideas about learning from precedents and exercises developed at MIT and ideas about physical description developed at Stanford. The strength of the synthesis is illustrated by way of representative experiments. All of these experiments have been performed with an implemented system.

Key Ideas

It is too hard to tell vision systems what things look like. It is easier to talk about purpose and what things are for. Consequently, we want vision systems to use functional descriptions to identify things, when necessary, and we want them to learn physical descriptions for themselves, when possible.

For example, there are many kinds of cups: some have handles, some do not; some have smooth cylindrical bodies, some are fluted; some are made of porcelain, others are styrofoam, and still others are metal. You could turn blue in the face describing all the physical possibilities. Functionally, however, all cups are things that are easy to drink from. Consequently, it is much easier to convey what cups are by saying what they are functionally.

To be more precise about what we are after, imagine that you are told cups are open vessels, standing stably, that you can lift. You see an object with a handle, an upward pointing concavity, and a flat bottom. You happen to know it is light. Because you already know something about bowls, bricks, and suitcases, you conclude that you are looking at a cup. You also create a physical model covering this particular cup type.

Our first purpose, then, is to explain how physical identification can be done using functional definitions. Our second purpose is to show how to learn physical models using functional definitions and specific acts of identification.

It is important to note that our theory of model learning involves a physical example and some precedents in addition to the functional definition:

- The physical example is essential, for otherwise there would be no way to know which precedents are relevant.
- The precedents are essential, for otherwise there would be no way to know which aspects of the physical example are relevant.

We now proceed to explain our function-to-form theory and to illustrate the ideas using some examples that have been run through our implementation.

Learning by Analogy and Constraint Transfer

We begin by reviewing the sort of tasks performed by the system that embodies the theory of learning by analogy and constraint transfer. For details, see Winston [1981].

- A natural language interface translates English sentences describing a precedent and a problem into links in a semantic net.

The input English interface was conceived and written by Katz. For details, see Katz and Winston [1982].

- A matcher determines a correspondence between the parts of the precedent and the problem. Figure 1 illustrates.
- An analogizer determines if the questioned link in the problem is supported by the given links in the problem. To do this, the analogizer transfers the CAUSE links supplied by the precedent onto the problem. Figure 2 illustrates.

†Patrick H. Winston and Boris Katz are at the MIT Artificial Intelligence Laboratory.

‡Thomas O. Binford and Michael Lowry are at the Stanford University Artificial Intelligence Laboratory.

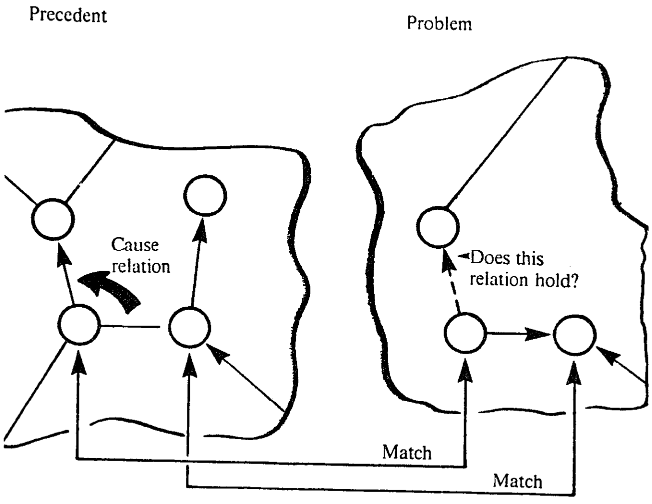

Figure 1. The matcher determines part correspondence using the links that populate the precedent and the problem. The matcher pays particular attention to links that are enmeshed in the CAUSE structure of the precedent.

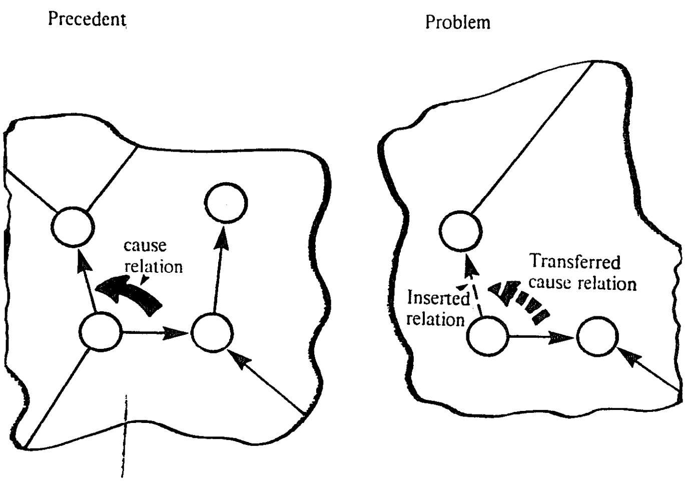

Figure 2. The analogizer transfers CAUSE constraints from the precedent to the problem. The problem is solved if links in the problem match the links carried along with the transferred CAUSE structure. In this simple illustration, there is only one CAUSE link. This CAUSE link leads from the link to be shown to a link that already holds.

- A rule generator constructs an if-then rule to capture that portion of the causal structure in the precedent that is ferreted out by the problem. The *then* part of the if-then rule comes from the questioned link in the problem. The *if* parts come from links identified by the transferred CAUSE structure.

Thus the rules look like this:

```
Rule
   RULE-1
if
    link found using CAUSE structure
    link found using CAUSE structure
    .
    .
    .
then
    link to be shown to hold
case
    names of all precedents used
```

In more complicated situations, no single precedent can supply enough causal structure. Consequently, several precedents must be strung together. A new precedent is sought whenever there is a path in the transferred CAUSE structure that does not lead to an already established link in the problem. The examples in this paper all use multiple precedents.

The Synthesis

In this section, we briefly describe the steps involved in our synthesis of the learning of ANALOGY and the physical representations of ACRONYM, an object modelling system based on generalized cylinders [Binford 1971, 1981, 1982; Brooks 1981]. In the next section we illustrate the steps by explaining a particular scenario in which the identification and learning system recognizes one physical model of a cup, expressed in ACRONYM's representation primitives, and then creates a model, in the form of an if-then rule, for that kind of cup. Here then are the steps in the synthesis:

1. Describe the thing to be recognized in functional terms.

The functional description is given in English and translated into semantic net links.

2. Show a physical example.

At the moment, the physical description is given in English, bypassing vision. The description is couched in the generalized-cylinder vocabulary of ACRONYM, however. Eventually this description will come optionally from ACRONYM.

3. Enhance the physical example's physical description.

The basic physical description, produced either from English or from an image, occasionally requires English enhancement. English enhancement is required when there is a need to record physical properties such as material composition, weight, and articulation, which are not easily obtained from a vision system or a vision-system prosthesis.

4. Show that the functional requirements are met by the enhanced physical description, thus identifying the object.

ANALOGY does this using precedents. Several precedents are usually necessary to show that all of the functional requirements are met.

5. Create a physical model of the functionally-defined concept.

ANALOGY creates a physical model in the form of an if-then rule whenever it successfully shows that a concept's functional requirements are met by a particular physical example. Since functional requirements usually can be met in a number of ways, there may be a number of physical models, each generated from a different physical example.

Once if-then based physical models are learned, examples of the concept can be recognized directly, without reference to functional requirements or to precedents. Moreover, ACRONYM can use the learned physical models to make top-down predictions about what things will be seen so that bottom-up procedures can look for those things.

Learning What an Ordinary Cup Looks Like

Now let us walk through the steps of the learning process, showing how to learn what an ordinary cup looks like.

The first step is to describe the cup concept in terms of functional qualities such as liftability, stability, and ability to serve as an accessible container. This is done by way of the following English:

> Let X be a definition. X is a definition of an object. The object is a cup because it is a stable liftable open-vessel. Remember X.

Of course, other, more elaborate definitions are possible, but this one seems to us to be good enough for the purpose of illustrating our learning theory.

The English is translated into the semantic net shown in figure 3.

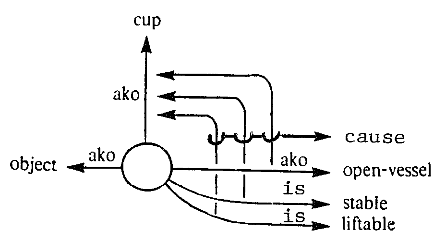

Figure 3. The functional definition of a cup. This semantic net is produced using an English description. AKO = A Kind Of.

The next step is to show an example of a cup, such as the one in figure 4. ACRONYM is capable of translating such visual information into the semantic net shown in figure 4. But inasmuch as our connection to ACRONYM is not complete, we currently bypass ACRONYM by using the following English instead.

> Let E be an exercise. E is an exercise about a red object. The object's body is small. Its bottom is flat. The object has a handle and an upward-pointing concavity.

In contrast to the definition, the qualities involved in the description of the particular cup are all physical qualities, not functional ones. (Assume that all qualities involving scales, like small size and light weight, are relative to the human body, by default, unless otherwise indicated.)

In the next step, we enhance the physical example's physical description. This enables us to specify physical properties and links that are not obtainable from vision.

> The object is light.

Now it is time to show that the functional requirements are met by the enhanced physical description. To do this requires using precedents relating the cup's functional descriptors to observed and stated physical descriptors. Three precedents are used. One indicates a way an object can be determined to be stable; another relates liftability to weight and having a handle; and still another explains what being an open-vessel means. All contain one thing that is irrelevant with respect to dealing with cups; these

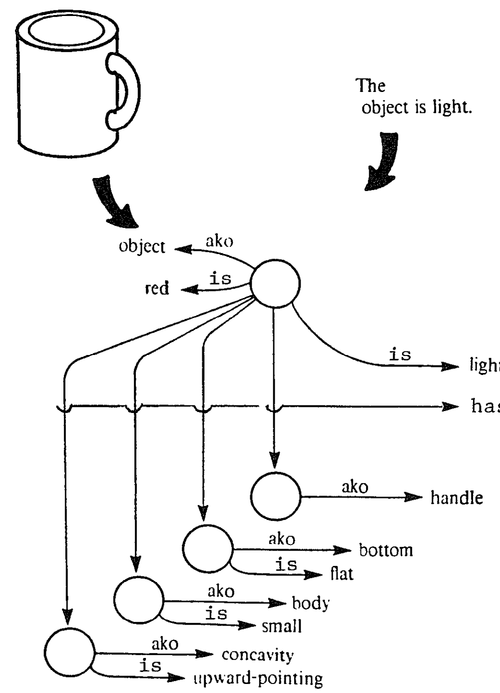

Figure 4. A particular cup, together with a semantic-net description of its physical appearance. Most of the semantic net can be produced by ACRONYM working from an image. Some nonvisual details must be produced by working through the natural language interface.

irrelevant things are representative of the detritus that can accompany the useful material.

> Let X be a description. X is a description of a brick. The brick is stable because its bottom is flat. The brick is hard. Remember X.
>
> Let X be a description. X is a description of a suitcase. The suitcase is liftable because it is graspable and because it is light. The

suitcase is graspable because it has a handle. The suitcase is useful because it is a portable container for clothes. Remember X.

Let X be a description. X is a description of a bowl. The bowl is an open-vessel because it has an upward-pointing concavity. The bowl contains tomato soup. Remember X.

With the functional definition in hand, together with relevant precedents, the analogy apparatus is ready to work as soon as it is stimulated by the following challenge:

In E, show that the object may be a cup.

This initiates a search for precedents relevant to showing something is a cup. The functional definition is retrieved. Next, a matcher determines the correspondence between parts of the exercise and the parts of the functional definition, a trivial task in this instance. Now the verifier overlays the cause links of the functional definition onto the exercise. Tracing through these overlayed cause links raises three questions: is the observed object stable, is it an open vessel, and is it liftable. All this is illustrated in figure 5.

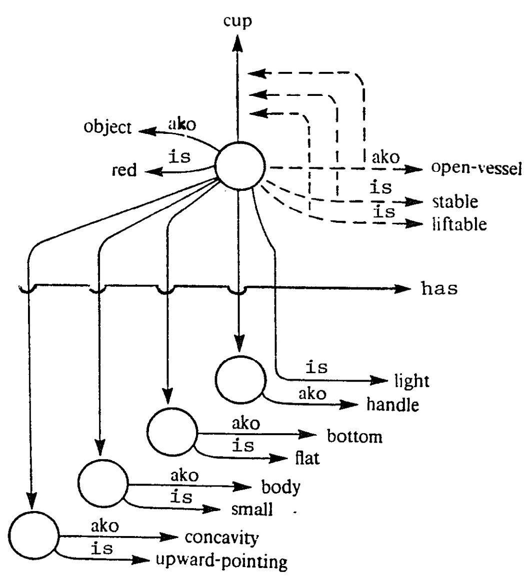

Figure 5. The cause links of a functional description, acting as a precedent, are overlayed on the exercise, leading to other questioned links. Note that all of the exercise description is physical, albeit not all visual. Overlayed structure is dashed.

Questioning if the object is liftable leads to a second search for a precedent, this time one that relates function to form, causing the suitcase description to be retrieved. The suitcase description, shown in figure 6, is matched to the exercise, its causal structure is overlayed on the exercise, and other questions are raised: is the observed object light and does it have a handle. Since it is light and does have a handle, the suitcase description suffices to deal with the liftable issue, leaving open the stability and open-vessel questions.

Thus the suitcase precedent, in effect, has a rule of inference buried in it, along with perhaps a lot of other useless things with respect to our purpose, including the statement about why the suitcase itself is useful. The job of analogy, then, is to find and exploit such implicit rules of inference.

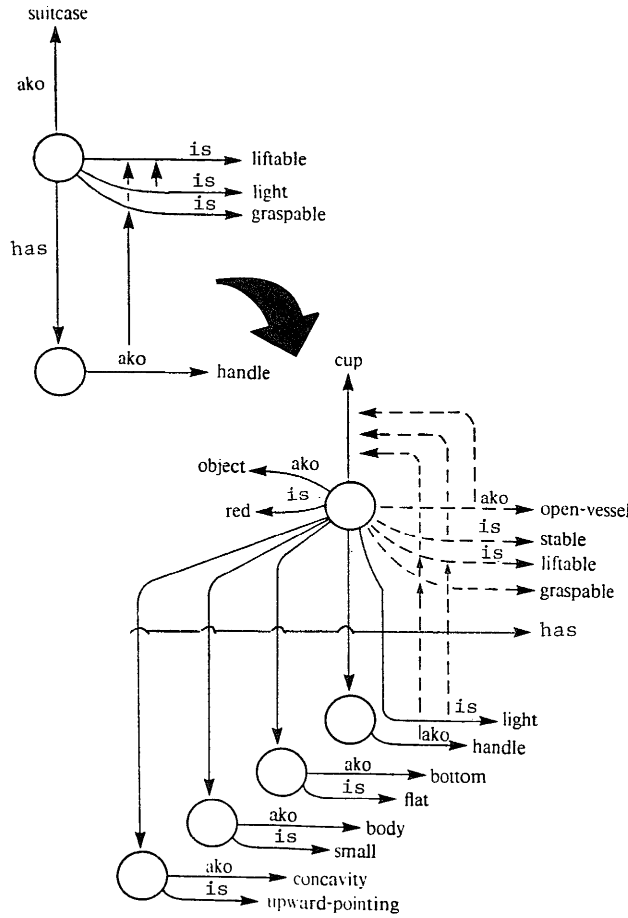

Figure 6. Cause links from the suitcase precedent are overlayed on the exercise, leading to questions about whether the object is light and whether the object has a handle. Overlayed structure is dashed. Many links of the suitcase precedent are not shown to avoid clutter on the diagram.

Checking out stability is done using the description of a brick. A brick is stable because it has a flat bottom. Similarly, to see if the object is an open-vessel, a bowl is used. A bowl is an open vessel because it has an upward-pointing concavity. Figure 7 illustrates.

At this point, there is supporting evidence for the conclusion that the exercise object is a cup.

Now we are ready for the final task, to build a physical model of the functionally defined concept. This is done by constructing an if-then rule from the links encountered in the problem-solving process: the questioned link goes to the *then* part; the links at the bottom of the transferred CAUSE structure go to the *if* part; and the intermediate

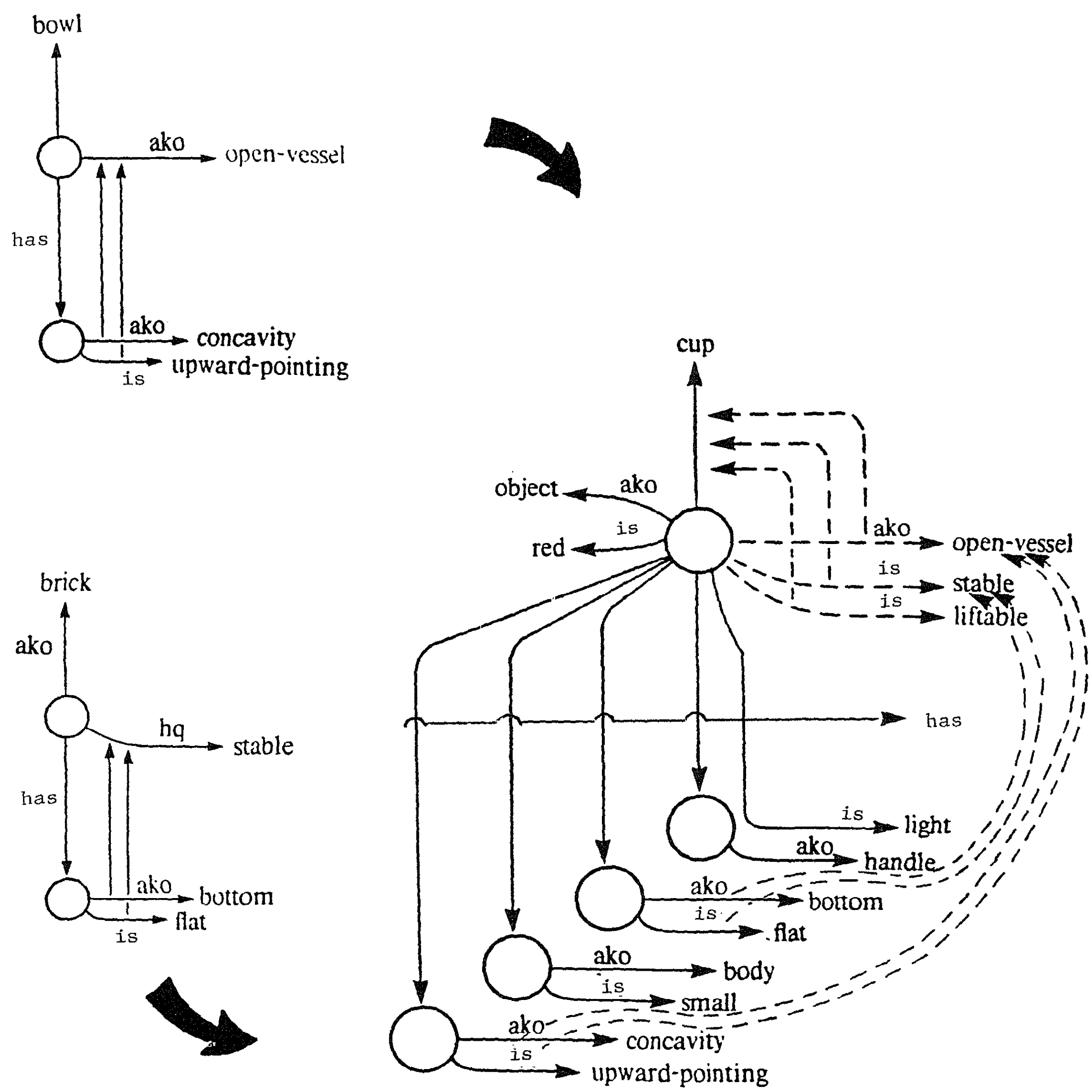

Figure 7. The brick precedent and the bowl precedent establish that the object is stable and that it is an open vessel. The cause links of the precedents are overlayed on the exercise, leading to questioned links that are immediately resolved by the facts. Overlayed structure is dashed. Many links of the precedents are not shown to avoid clutter on the diagram.

links of the transferred CAUSE structure go into the *unless* part.[1]

[1]The *unless* parts come from the links lying between those links supplying the if and then parts of the rule. For a rule to apply, it must be that there is no direct reason to believe any link in the rule's *unless* part, as explained in an earlier paper [Winston 1982].

The result follows:

```
Rule
    RULE-1
if
    [OBJECT-9 IS LIGHT]
    [OBJECT-9 HAS CONCAVITY-7]
    [OBJECT-9 HAS HANDLE-4]
    [OBJECT-9 HAS BOTTOM-7]
    [CONCAVITY-7 AKO CONCAVITY]
    [CONCAVITY-7 IS UPWARD-POINTING]
    [HANDLE-4 AKO HANDLE]
    [BOTTOM-7 AKO BOTTOM]
    [BOTTOM-7 IS FLAT]
then
    [OBJECT-9 AKO CUP]
unless
    [[OBJECT-9 AKO OPEN-VESSEL] IS FALSE]
    [[OBJECT-9 IS LIFTABLE] IS FALSE]
    [[OBJECT-9 IS GRASPABLE] IS FALSE]
    [[OBJECT-9 IS STABLE] IS FALSE]
case
    DEFINITION-1 DESCRIPTION-2 DESCRIPTION-3 DESCRIPTION-1
```

At first the unless conditions may seem strange, for if all the ordinary conditions hold and if the causal connections in the precedents represent certainties, then none of the unless conditions could trigger. However, the learning system assumes that the precedents' causal connections indicate tendencies, rather than certainties. Consequently, from the learning system's perspective, the unless conditions must appear. An earlier paper gives several examples where similar unless conditions are necessary [Winston 1982].

Learning What a Styrofoam Cup Looks Like

Styrofoam cups without handles are described by another rule that is learned in the same way using the same functional description. The only differences is that liftability is handled by way of a flashlight precedent rather than by the suitcase precedent.

> Let X be a description. X is a description of a flashlight. The flashlight is liftable because its body is graspable and because the flashlight is light. The flashlight's body is graspable because it is small and cylindrical. Remember X.

Thus the learned rule is as follows:

```
Rule
    RULE-2
if
    [OBJECT-10 IS LIGHT]
    [OBJECT-10 HAS BODY-9]
    [OBJECT-10 HAS CONCAVITY-8]
    [OBJECT-10 HAS BOTTOM-8]
    [CONCAVITY-8 AKO CONCAVITY]
    [CONCAVITY-8 IS UPWARD-POINTING]
    [BODY-9 AKO BODY]
    [BODY-9 IS CYLINDRICAL]
    [BODY-9 IS SMALL]
    [BOTTOM-8 AKO BOTTOM]
    [BOTTOM-8 IS FLAT]
then
    [OBJECT-10 AKO CUP]
unless
    [[OBJECT-10 AKO OPEN-VESSEL] IS FALSE]
    [[OBJECT-10 IS LIFTABLE] IS FALSE]
    [[OBJECT-10 IS STABLE] IS FALSE]
    [[BODY-9 IS GRASPABLE] IS FALSE]
case
    DEFINITION-1 DESCRIPTION-2 DESCRIPTION-4 DESCRIPTION-1
```

Recognizing Cups and Using Censors

We now have two descriptions that enable direct recognition of cups. These can be used, for example, on the following descriptions, conveyed by ACRONYM or by the natural language interface:

> Let E be an exercise. E is an exercise about a light object. The object's body is small. The object has a handle. The object's bottom is flat. Its concavity is upward-pointing. Its contents are hot. In E show that the object may be a cup.
>
> Let E be an exercise. E is an exercise about a light object. The object's bottom is flat. Its body is small and cylindrical. Its concavity is upward-pointing. Its contents are hot. Its body's material is an insulator. In E show that the object may be a cup.

For the first of these two exercises, the rule requiring a handle works immediately. It is immaterial that the contents of the cup are hot.

For the second, the rule requiring a small, cylindrical body works immediately. Again it is immaterial that the contents of the cup are hot since nothing is known about the links among content temperature, graspability, and insulating materials. Proving some knowledge about these things by way of some censors makes identification more interesting.

Suppose, for example, that we teach or tell the machine that an object with hot contents will not have a graspable body, given no reason to doubt that the object's body is hot. Further suppose that we teach or tell the machine that an object's body is not hot, even if its contents are, if the body is made from an insulator. All this is captured by the following censor rules, each of which can make a simple physical deduction:

> Let C1 be a Censor. C1 is a censor about an object. The object's body is not graspable because its contents are hot unless its body is not hot. Make C1 a censor using the object's body is not graspable.
>
> Let C2 be a censor. C2 is a censor about an object. The object's contents are hot. Its body is not hot because its body's material is an insulator. Make C2 a censor using the object's body is not hot.

Repeating the second exercise now evokes the following scenario:

Asking whether the object is a cup activates the rule about cups without handles. The *if* conditions of the rule are satisfied.

The *unless* conditions of the rule are checked. One of these conditions states that the object's body must not be plainly ungraspable.

Asking about graspability activates the censor relating graspability to hot contents. The censor's *if* condition is satisfied, and the censor is about to block the cup-identifying rule. The censor's *unless* condition must be checked first, however.

The censor's *unless* condition pertains to hot bodies. This condition activates a second censor, the one denying that a body is hot if it is made of an insulator. This second censor's *if* condition is satisfied, and there are no *unless* conditions.

The second censor blocks the first censor. The first censor therefore cannot block the cup-identifying rule. The rule identifies the object as a cup. It would not have worked if the contents were hot and the body were made from something other than an insulator.

Related Work

The theory explained in this paper builds directly on two sets of ideas: one set that involves a theory of precedent-driven learning using *constraint transfer* [Winston 1979, 1981, 1982]; and another set that involves model-driven recognition using *generalized cylinders* [Binford 1971, 1981, 1982; Brooks 1981].

Another important precedent to this paper is the work of Freeman and Newell on the role of functional reasoning in design. In their paper on the subject [1971], they proposed that available structures should be described in terms of functions provided and functions performed, and they hinted that some of this knowledge might be accumulated through experience.

Another way to learn what things look like is by near misses. Since the use of near misses was introduced by Winston [1970], several researchers have offered improved methods for exploiting near miss information. See Dietterich and Michalski [1981] for an excellent review of work by the authors, Hayes-Roth, and Vere. Also see work by Mitchell [1982].

Acknowledgments

This paper was improved by comments from Robert Berwick, Randall Davis, and Karen A. Prendergast.

References

Binford, Thomas O., "Visual Perception by Computer," *Proc. IEEE Conf. on Systems Science and Cybernetics*, Miami, December, 1971.

Binford, Thomas O., "Inferring Surfaces from Images," *Artificial Intelligence*, vol. 17, August, 1981. Volume 17 is also available as *Computer Vision*, edited by J. Michael Brady, North-Holland, Amsterdam, 1981.

Binford, Thomas O., "Survey of Model-Based Image Analysis Systems," *Robotics Research*, vol. 1, no. 1, Spring, 1982.

Brooks, Rodney A., "Symbolic Reasoning Among 3-D Models and 2-D Images," PhD Thesis, Stanford University, Computer Science Department report STAN-CS-81-861, 1981. A shorter version is "Symbolic Reasoning Among 3-D Models and 2-D Images," *Artificial Intelligence*, vol. 17, August, 1981. Volume 17 is also available as *Computer Vision*, edited by J. Michael Brady, North-Holland, Amsterdam, 1981.

Dietterich, Thomas G. and Ryszard S. Michalski., "Inductive Learning of Structural Descriptions," *Artificial Intelligence*, vol. 16, no. 3, 1981.

Freeman, P, and Allen Newell, "A Model for Functional Reasoning in Design," *Proceedings of the Second International Joint Conference on Artificial Intelligence*, London, England, 1971.

Katz, Boris and Patrick H. Winston, "Parsing and Generating English using Commutative Transformations," Artificial Intelligence Laboratory Memo No. 677, May, 1982. A version is also available as "A Two-way Natural Language Interface," in *Integrated Interactive Computing Systems*, edited by P. Degano and E. Sandewall, North-Holland, Amsterdam, 1982. Also see "A Three-Step Procedure for Language Generation," by Boris Katz, Artificial Intelligence Laboratory Memo No. 599, December, 1980.

Mitchell, Tom M., "Generalization as Search," *Artificial Intelligence*, vol. 18, no. 2, 1982.

Winston, Patrick Henry, "Learning Structural Descriptions from Examples," Ph.D. 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 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," M.I.T. Artificial Intelligence Laboratory Memo No. 520, April 1979.

Winston, Patrick Henry, "Learning New Principles from Precedents and Exercises," *Artificial Intelligence*, vol. 19, no. 3. A version with details is available as "Learning New Principles from Precedents and Exercises: the Details," M.I.T. Artificial Intelligence Laboratory Memo No. 632, May 1981.

Winston, Patrick Henry, "Learning by Augmenting Rules and Accumulating Censors," M.I.T. Artificial Intelligence Laboratory Memo No. 678, May 1982.

This research was done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for MIT'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.