2

Images and Inference*

ROBERT K. LINDSAY
University of Michigan

Abstract

It is frequently asked whether imagery differs in a fundamental way from other forms of knowledge representation, specifically the predicative forms employed in artificial intelligence programs. Frequently suggested distinctions are pictorial versus descriptional, and analog versus digital. This paper argues that these distinctions are not central in understanding the role of imagery in cognition, and moreover do not correctly capture the difference between visual perception and language. A distinction is proposed between the representation of images, on the one hand, and a calculus-plus-proof-procedure form of knowledge representation on the other. This distinction is not based upon differences in expressive power or form, but rather is based upon a distinction between how these two representations function, specifically how they are used to make inferences. On this view, an important functional role of imagery is to provide a non-proof-procedural method for inference, using a constraint satisfaction mechanism. Images, even the limited class of images here called diagrams, support inference in a way that is distinct from the way predicative representations support inference. This analysis offers an approach to solving the "frame problem" of cognitive science.

1. Introduction

The notion that mental imagery plays an important and distinct role in cognition has a long history. Indeed, there was once an extended debate as to whether *any* thought was possible without *imagery* (Woodworth, 1938). Today the pendulum has swung the other way (to invoke an image-based metaphor), and there is debate as to whether imagery, *as a distinct form of representation*, plays *any* role in cognition (Dennett, 1981a, 1981b; Fodor,

1981; Kosslyn, 1975, 1980; Kosslyn, Pinker, Smith, & Schwartz, 1979; Kosslyn & Pomerantz, 1977; Pylyshyn, 1980; Shepard, 1973, 1978).

Advocates of a distinct role of imagery in cognition have emphasized a variety of funtions that imagery might serve. These include memory (Bower, 1972; Paivio, 1971), navigation (Evans, 1980), and perceptual recognition (Posner & Keele, 1967). A recent review of empirical studies of mental imagery and its functional similarities to visual perception may be found in Finke and Shepard (1986). In this paper I focus on imagery's function in inference. Inference is an important component of most if not all other cognitive processes, including perception (Rock, 1983; but see also Gibson, 1979), language (Lindsay, 1963), navigation (Gladwin, 1970), memory (Bartlett, 1932; Chase & Simon, 1973) and problem solving (Larkin & Simon, 1987).

Any cognitive system, natural or artificial, must draw inferences from its knowledge. If its environment is changing, as would be the normal case, the inference problem becomes acute, because a change in even a single item of knowledge might have widespread effects on many others. The problem of determining which items are affected and how they are affected has been called the "frame problem" in artificial intelligence (Haugeland, 1985; McCarthy & Hayes, 1969; Raphael, 1971). The inferences in question need not involve long chains of deduction (although they may); the frame problem is difficult because of the sheer number of inferences possible in the face of large quantities of knowledge whose mutual dependencies must somehow be specified or derived. Obviously the frame problem is one to which any theory of human cognition must supply an answer.

The major theme of research attacking the frame problem has been the search for appropriate modifications and extensions of first-order predicate logic (FOL) that will permit the description of world knowledge and situations and the formulation of rules of deduction that describe how situations change over time (a so-called situational calculus). One requirement of such a calculus is that conclusions must be retractable in the face of additional information; that is, unlike FOL where conclusions only accumulate, the logic must be "non-monotonic." There is an extensive literature on these topics (for a review, see Reiter, 1987).

In this paper I propose a different approach to the frame problem in which images are employed as inference-making representations. Specifically, I propose a model for representing geometric diagrams in a way that permits the drawing of inferences without the explicit use of rules of deduction. The kernel of this idea was present in an early artificial intelligence program (Lindsay, 1961, 1963) which extracted kinship facts from linguistic inputs and used a family tree representation to make explicit some inferences that were only implicit in the inputs. Visual imagery, in addition to whatever other

characteristics it has, provides a much more general and ubiquitous application of this idea. Haugeland (1985, p. 229), in commenting on the frame problem and imagery, puts the essential idea succinctly: "The beauty of images is that (spatial) side effects take care of themselves." Here I attempt to spell out for the case of geometric diagrams (which illustrates the more complex case of spatial imagery in general) just what is required for the "side effects" (inferences) to "take care of themselves."

One might well ask why inference is not adequately accounted for by predicative knowledge representations employing logic, since inference is a task for which logic was explicitly devised and for which there is a well-developed theory. My answer is that visual images possess properties (to be described shortly) not possessed by *deductive* propositional representation, and these properties help avoid the combinatorial explosion of correct but trivial inferences that must be explicitly represented in a propositional system. Accordingly, an important role of imagery in cognition is as a constructive inferential knowledge representation system that efficiently makes inferences based on one's beliefs (including one's accurate knowledge) of how real-world situations behave. It is one's beliefs, perhaps describable as predicative knowledge, that are used to construct images; inferences are retrieved from these images. The crucial property that distinguishes images from other knowledge representations is that they are non-deductive, that is non-proof-procedure based. In order to explain this view it is necessary to take a more careful look at what a knowledge representation system is.

2. Knowledge representation systems

Although a general theory of knowledge representation is not at hand, the past two decades have seen a large number of computer implementations of methods for storing and processing knowledge. From this work has emerged certain limited generalizations in the form of programming systems for expressing procedural and factual knowledge of various sorts, and several authors (e.g., Bobrow, 1975; Palmer, 1978; see also Brachman & Levesque, 1985) have offered analyses of the representation problem that have clarified several important issues.

To avoid the endless regress of the homunculus fallacy, it is now generally recognized that it is necessary to specify not only the structure of the representation but also how it is constructed, and how it is accessed and used. A completely specified knowledge representation system therefore consists of at least three parts: a set of *construction processes*, a set of *representation structures*, and a set of *retrieval processes*. A specification of the form of the

representation structures alone has no explanatory power; however, a representation system with unspecified construction processes could still model those functions that use *externally* provided representations. The work here presented, however, addresses issues of the selective construction of *internal* representations and how construction and retrieval processes interact. Thus a full model of imagery-based abilities must prescribe all three components of the representation system.

A knowledge representation system should be described in the context of a task or purpose that it will serve; different tasks or purposes often, though perhaps not always, are better served by different representation systems. Thus even if a single representation system can in principle serve all purposes, it often serves one of them *better* (e.g., Amarel, 1968). Since a knowledge representation system is not just a passive repository of facts, but includes methods for applying and retrieving information as well, issues of efficiency arise and may be criterial for judging theoretical adequacy. However, I am making no claim that the proposed representation system for images can make inferences that are *in principle* beyond, say, predicate logic representation systems.

Any knowledge representation system must record some information, that is, make it available at a later time. This may be thought of as a special, limited form of inference even if what is available at a later time is exactly what was entered explicitly. (The frame problem is non-trivial simply because it is not always correct to conclude that facts do not change over time, hence static memory often makes incorrect "inferences.") However, I wish to reserve the term *inference* for its more customary use: *making explicit* information that was *implicit* in sets of inputs (Lindsay 1961). Thus I will restrict the term *inference-making knowledge representation systems* (IKRs) to representation systems whose three components jointly yield information that was not directly provided to it. For example, a device that translates English text into predicate logic formulas and then back to English would be a knowledge representation system, but not an IKR. If it *also* applied a proof procedure to the formulas to generate other formulas, which could then be translated into English, it would be an IKR.

A knowledge representation system that did English–logic and logic–English translation and was able to retrieve inferences from inputs, but *did not embody a separate proof procedure* may seem paradoxical. If the representation structure is FOL, that is indeed the case, for logic *requires* a proof procedure for making inferences; that is built into the concept of logic. However, there can be knowledge representation systems that make inferences without the use of explicit rules of deduction but simply by virtue of the properties of the knowledge representation system alone.

To illustrate this, consider a simple case consisting of a discrete grid, each cell of which could be occupied by a single point labelled by a letter. Consider the input *b is one grid point due right of a, c is directly above b, and d is one grid point due left of c*. From this we may conclude that *d is directly above a*. This inference could be supported by a calculus based system with appropriate rules of deduction; such a system would be *deductive*. Alternatively, the inference would be supported by a system of construction and retrieval routines that placed letters on the described grid and read off relationships by scanning the grid; such a system would be *non-deductive*. The non-deductive system requires no separate computational inference-making stage: the operation of the construction process entails the "making" of the inferences.

Similar tasks have been studied experimentally, attempting to establish that human subjects employ visual imagery in their performance. Huttenlocher (1968) and others have examined performance in solving problems requiring inferences from statements such as *Tom is taller than Sam* and *John is shorter than Sam*; the evidence (from errors and response times) supports the hypothesis that subjects image tokens of Tom, and so forth, translating height relations into spatial relations in the image. In this example, the imaging processes may be construed as "implementing" certain properties of a *simple ordering* relation (transitivity, anti-symmetry, and irreflexivity); Elliott (1965) provides examples of other classes of relations that can be treated with similar methods. Clearly, such a knowledge representation can make some inferences, and yet the inferences inhere in the knowledge representation, including the construction and retrieval processes, and do not require a separate proof procedure. The construction and retrieval processes are exactly the same whether the retrieved knowledge was given explicitly or inferred.

These examples are illustrations of non-proof-procedure methods of inference that represent a substantially different approach to inference from the standard view of logic, and that may provide a connection between inference and mind that does not rely on computationally opaque and inefficient methods. The suggestion is not without precedent. Lindsay (1963) proposed "inferential memory" as a non-rule-based form of deduction. Quillian's (1968) "semantic nets" and later elaborations permit non-proof-procedural deduction. Clearly "inheritance of properties" in semantic nets (Brachman, 1979; Fahlman, 1979, 1984) employs such a method. However, a more fundamental and general connection to cognition exists: imagery.

The basic hypothesis of this paper is that *visual imagery employs non-proof-procedural knowledge representations that support inference by a constraint satisfaction mechanism built into the processes that construct and access them*. A representation that possesses this quality may remain symbolic, and even digital, but is not based on predicate logic. This hypothesis is elaborated

below for computer-based knowledge representation; I suggest that it is also true for human (mental) knowledge representation, though at present I can offer no empirical support for that hypothesis.

Non-deductive inference should not be *equated* with visual imagery, for then some "images" would have no specific spatial properties. For example, inheritance of properties in semantic nets is non-deductive but not explicitly spatial. On the other hand, there is a natural translation of such inheritance (which is based on the set-inclusion relation) into spatial terms, namely Euler/Venn diagrams. Indeed such translations are customary and powerful methods of problem solving by analogy and frequently are suggested as heuristic devices in mathematical texts. Haugeland (1985) attributes such methods, perhaps overgenerously, to Galileo, who used spatial metaphors to prove results concerning non-spatial propositions. Perhaps it is the case that all non-deductive representations that are available to human thought could be translated into spatial analogies. This is not a claim, however, that I am prepared to make, and thus leave open the possibility that imagery does not tap all such methods.

Conversely I do not claim that non-deductive inference is *all* there is to imagery. Finke and Shepard (1986, p. 37) conclude from their recent review of experimental studies of imagery that "... the functional equivalence hypothesis [that the internal processes are essentially the same in perceptual and imagery tasks] provides the *single best* overall explanation for the results reviewed in the several sections." Many aspects of the percept-like character of imagery (e.g., color, texture, depth, and so forth) are not addressed in the model to be presented, and some of them may serve important functions other than inference (e.g., simple recall, emotional impact, aesthetic judgment, and so forth).

3. A knowledge representation for diagrams

For purposes of illustration, in the remaining discussion the category of visual images will be restricted to *diagrams*: informally, these are drawings that can be made with paper and pencil with a straight-edge and compass, but without color or continuous grey-scale shading. A more precise definition of diagrams is given below (in the form of construction processes). Basically, diagrams include those things that can be drawn on a black and white computer terminal by "graphics software." Diagrams *are* expressible as a set of propositions, as witnessed by their representation with such software. However, such propositional representations are not in general perspicuous, for reasons later addressed.

I now present a more precise model of non-deductive inference (NDI), restricted to that subset of diagrams composed of points, straight lines, circular arcs, and symbolic labels for these components. The omission of many important qualities of actual drawings, such as colors, textures, widths of lines, and so forth, may well limit the range of inferences that can be supported by this model, but what remains to be addressed is an important and ubiquitous set of mental activities. The following are the basic functions of the proposed representation for these two-dimensional diagrams.

Representation structure

The elements of a representation structure are symbolic names for points, line segments, and circular arcs, combined into expressions that relate these symbols to their locations on a two-dimensional, bounded flat *tablet*. A representation structure R for a diagram D is a specific set of such symbols and expressions.

Four things must be kept distinct: (1) the diagram which is represented, (2) that which the diagram denotes, (3) the representation structure for the diagram, and (4) the class of potential representation structures for diagrams in general. For example, we might wish to represent a specific diagram, such as a floor plan (1) of the White House. That floor plan denotes the layout of the building at 1600 Pennsylvania Avenue (2) in Washington. The representation structure (3) is a set of coordinates of pairs of tablet points (corresponding to the end points of the line segments in the floor plan) plus names for each of these points, plus several sets of points (corresponding to a sample of points on each line segment and curve of the floor plan), plus names of these line segments and curves. All points that are part of the representation structure are "marked," that is, they are distinguished from the other points on the tablet.

The class of representation structures (4) could be defined by a grammar specifying the set of real numbers that may serve as coordinates on the tablet, the classes of coordinate combinations (e.g., pairs) that form permissible expressions, what symbols may serve as labels, and so forth. In fact, however, the usual forms of immediate constituent rule grammars are not perspicuous descriptions of the non-linear structures here employed. Consequently, the class of representation structures will merely be specified implicitly by the construction processes themselves. These processes implicitly obey a grammar of representations, but are not identical to it, just as a program that generates English might obey a grammar of English, but not *be* a grammar of English itself.

The implicit grammar of the representation structure defines specific clas-

ses of objects (points, lines, circles, etc.) that consist of sets of coordinates that lie within the bounded rectangular tablet. The tablet points are dense in the usual sense that, for any two given points reference can be made to a point between them. Points on the tablet may be labelled with symbolic names, and may be either *marked* or *unmarked*.

Construction processes

The set of construction processes is formed from the following *primitive construction processes*:

Mark a given point on the tablet.
Erase a given point; that is, unmark it.
Label a given point with a specified name.
Construct due right/left point relative to a given point.
Construct due above/below point relative to a given point.
Construct a line segment between two given points (i.e., mark a covering set of points between the end points).
Select the midpoint on a given line segment and return its coordinates.
Construct a perpendicular to a given line segment (perhaps extended) from a given point (i.e., mark a covering set of the perpendicular).
Construct a parallel to a given line segment through a given point.
Parallel-translate a given line segment (i.e., mark its extension along itself) until a given *condition predicate* (see below) is true.
Pass a circle through three given points (i.e., mark a covering set of the circle).
Construct a circle with a given point as center and passing through a second given point (i.e., mark a covering set of the circle).
Select a point on the tablet meeting a specified ordered list of *condition predicates* (see below).
Extend a given line segment until a given ordered list of *condition predicates* is met.
Perpendicular-translate a line segment until a given ordered list of *condition predicates* is met.
Rotate a given line segment about a given point until a given ordered list of *condition predicates* is met (i.e., unmark its covering set and mark the covering set of the line segment that would result after a rotation meeting the condition).

In the above construction processes, *condition predicates* are statements that are either true or false and can be verified by compositions of the retrieval processes, enumerated below.

Retrieval processes

The set of retrieval processes are formed from the following *primitive retrieval processes*:

Right/left order: Given two points determine which one is rightmost or that neither is.

Above/below order: Given two points determine which one is above or that neither is.

Point–line relation: Determine if a given point is on a given line segment, and if not determine which side of the line it is on.

Point–circle relation: Determine if a given point is inside, on, or outside a given circle.

Closure: Determine if a given set of line segments form a continuous, closed curve; this process returns *true* or *false*.

Comparative length: Determine which of two given line segments is longer, or whether they are of equal length within stated resolution; returns the longer segment, or *false* if they are of equal length.

Comparative angle: Determine which of two given angles is greater, or whether they are equal to within stated resolution.

Intersection: Determine if and where two given line segments intersect; returns the point of intersection, or *false*.

Angle metric quality: Determine if two given line segments are parallel and if not determine if the (directed) angle from the first to the second is zero, acute, right, obtuse, straight, oblique, a negative right angle, or reflex; returns *zero*, *acute*, *right*, *oblique*, and so forth.

Mirror symmetry: Determine if the entire representation structure exhibits mirror symmetry about a given line.

Some of the retrieval processes are predicative statements; those that are not can be converted into one or more predicative statements. For example, *closure* can be used directly as a condition predicate, and *angle metric quality* yields eight condition predicates. Similarly, the composite retrieval processes, such as those illustrated below, may yield additional condition predicates.

The condition predicates are the heart of the inferential power of the knowledge representation system. Consequently, their efficient implementation is extremely important to the problem solving power of the model. In this paper, however, I discuss only their functional requirements and present what is hoped to be a *sufficient* set of processes to accomplish the purposes described in section 4. It is not claimed that the set is unique, nor that the processes correspond in any direct way to neural or mental processes of humans.

From the primitive construction and retrieval processes, more complex construction and retrieval processes can be formed. For example:

Line–circle relation: Determine which points on a given line segment are outside, on, or inside the circle; returns three lists of coordinates representing the endpoints of these sets of points. (Uses point–circle relation.)

Parity: Determine if two given points are on the same or opposite sides of a given line segment (extended), or neither (that is, one or both lie on the line segment); returns *same, opposite*, or *false*. (Uses point–line relation).

Colinearity: Determine if three given points lie on the same line. (Uses point–line relation and parallel–translate).

Construct rectangle: Given two parallel line segments, move one along itself until the two line segments form opposite sides of a rectangle. (Uses parallel–translate, construct–perpendicular, comparative length, and angle metric quality).

Due right: Given two points, determine if the first is due right of the second. (Uses right/left order and above/below order).

Note that many of these processes have no knowledge of metric properties; exceptions are midpoint construction, the two circle constructions, comparative length, comparative angle, and angle metric quality. The metric information in these processes, however, is "qualitative" (in the sense of de Kleer & Brown, 1984 and Forbus, 1984); the distinct values are limited to a small, ordered set, and distance measures are only comparative. These qualitative metric properties are necessary for dealing with direction and relative length, and these suffice for a wide class of inferences, though not all conceivable ones that might be supported by diagrams.

With this knowledge representation system it is possible to have non-deductive inference. The construction processes employ methods that impose constraints on what can be constructed, and these constraints preclude the construction of representation structures that embody (make available to the retrieval processes) invalid inferences. In some cases, enough invalid inferences are eliminated to leave some unambiguous, true conclusions.

No claim is made that NDI methods can in principle do things that logic cannot. It is likely, however, that they can do some things *more efficiently*, and that is the crux of the matter. The efficiency edge need not always be on the side of NDI knowledge representations, and yet humans, and machines, may use them more widely than would be prudent simply because the machinery is already in place, just as one might pound in screws with a nearby hammer rather than fetching a distant screwdriver. Note that this model of imagery is indeed computational since it may be implemented on a digital

computer with standard (serial, digital) computational techniques. However, to achieve an efficiency advantage may require the use of non-conventional computers, including analog and parallel methods. Regardless of implementation, the model differs conceptually from proof-procedure and production-rule based inference schemes. The next section illustrates how the sorts of inference mechanism this method provides may be used in geometric problem solving and thinking.

4. Uses of NDI knowledge representation in cognition

There is, of course, more to cognition than the making of inferences. In this section I discuss how diagrammatic representations can be used in the service of goal-directed processes such as theorem proving and problem solving.

In the Geometry Machine (Gelernter, 1959), an early artificial intelligence program, the inferences were made by a calculus that generated a problem-subproblem tree and sought to discover a non-rigorous but formal deductive proof of a given theorem. An important feature of this program was its use of a diagram as a mechanism for rejecting implausible subproblems. For example, if the calculus proposed proving that two triangles were congruent, the program first consulted its diagram to see if these two triangles had the same perimeter length; if not the subgoal was abandoned. Here the diagram is of service precisely because it is not possible to construct a diagram that depicts two triangles of unequal perimeter while remaining consistent with premises that imply that the triangles are congruent. It is of course possible for the diagram to err in the other direction: the premises may not imply congruency, but the chosen diagram may represent equal perimeter triangles, since geometry diagrams are in general under-determined. The Geometry Machine strategy is thus conservative: it may not reject all false paths, but those paths rejected are indeed false. Gelernter, Hansen, and Loveland (1960) estimated from their experiments that the amount of search was reduced by a factor of at least 200 by this method.

It is interesting to note that the use of the diagram as heuristic device in this way does not require that a sketch be "shown" to the computer. In fact, that would not work since the Geometry Machine did not have any vision programs. Instead, the functional equivalent of a diagram is required. In actuality, the programmers supplied "diagrams" that consisted of a list of numerical coordinates of the points referred to by the premises. Nonetheless, from our viewpoint, the list of coordinates functioned diagrammatically for the computer.

The Geometry Machine used diagrams to *dispose of* conjectures. Diagrams

may also be used to *propose* conjectures. For example, attempting to draw diagrams consistent with a set of premises may force the construction of, say, an isosceles triangle rather than a scalene triangle; this may be taken to suggest that the forced feature follows from the premises. The Geometry Machine could not do this since it did not construct diagrams, but Lenat's (1976) AM program used essentially this method of conjecturing (but not with diagrams) when its generation of examples turned up "interesting" features.

Similarly, an NDI model does not always eliminate search, but aids search by the use of constraint satisfaction heuristics that can either propose or dispose of putative inferences.

Next consider Figure 1. *ABC* is a right triangle; a perpendicular from the right angle to the opposite side has been constructed, forming the two triangles *DAB* and *DCA*.

Inference 1: Area of *ABC* = area of *DAB* + area of *DCA*.
Inference 2: DAB is similar to *DCA*.

Inference 1 is direct, immediate, and compelling simply from an examination of the diagram; Inference 2 requires study but nonetheless results in conviction that its proof (in a formal calculus) could be readily obtained. How can the diagram serve these functions, and why are such conclusions more "direct" than the formal textbook proofs employing propositional statements?

Inference 1 of course is an instance of the whole being equal to the sum of its parts. The conclusion does not depend on the fact that *ABC* is a triangle, or even that the diagram consists solely of linear elements; an arbitrary shape containing a dividing contour would yield the analogous conclusion. A pro-

Figure 1.
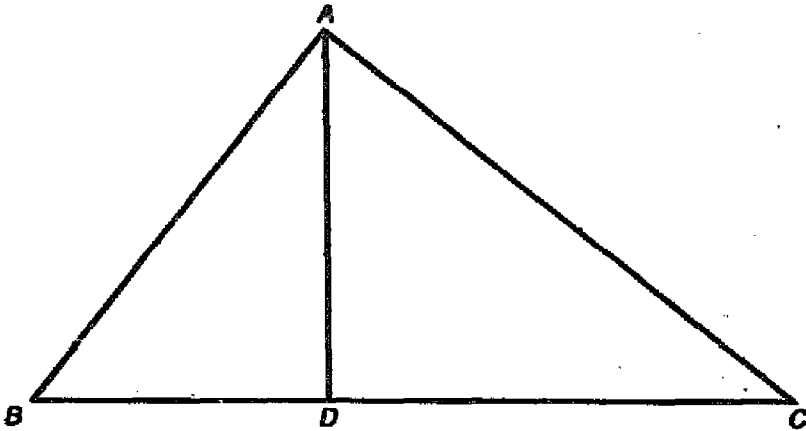
cess that spreads "activation" through bounded areas (such as the "coloring" method proposed by Ullman, 1984) could discover the validity of Inference 1.

On the other hand, proving Inference 1 deductively for the general case is a complicated procedure. Figure 1 is a planar graph (Harary, 1969) that can be specified propositionally by listing the nodes (A, B, C, D) and the edges (AB, BD, DC, CA, AD). From this specification alone it is possible to show that the graph has exactly three *meshes*[1] – *ABDA*, *DCAD*, and *ABDCA* – of which only *ABDCA* would remain if the edge *AD* were removed. This is the non-metric analog of the addition of areas property, yet, to extract the mesh structure from the list of nodes and edges by serial computation is neither straightforward nor perspicuous. The algorithm of Tarjan (1971) for (essentially) this task, which he proved to be within a constant factor of optimal in efficiency, is expressed as some 500 lines of high-level code and more than 130 pages of explanation.[2]

In contrast to Inference 1, Inference 2 does depend on certain metric properties of the diagram, such as the recognition of right angles and the ratios of lengths of sides. For the human observer, the precision of the draw-ing does not need to be great. As noted earlier, the metric precision in the specification of our construction and retrieval processes has been limited to "qualitative" values: they need to be able to distinguish right angles from non-right angles, identify equivalent lengths of line segments, and make a few other distinctions. This seems to be appropriate as a model of human use of metric information here and in other problem solving tasks (de Kleer & Brown, 1984). A slight sloppiness in drawing the perpendicular *AD* so that the angle *BDA* is, say, 87°, does not interfere with a human's use of a diag-ram, and a similar coarseness of measurement is all that is available in our model. Similarly, a person does not need to know the actual values of the lengths of segments, or even of their ratios, they merely must "look" appro-priate. We notice "by inspection," and our model would "notice" by process *comparative length*, that the order of the sides of *DCA* by length is $AD <$ $DC < CA$, and the order of the sides of *DAB* by length is $BD < DA < AB$. We can then notice "by inspection," and our model would "notice" by sys-tematic search, that the corresponding pairs of sides have the same ordering by length: *AD* is longer than *BD*, *DC* is longer than *DA*, and *CA* longer than *AB*. Why are these the things that we focus on, rather than any of a number

---

[1]A *mesh* of a planar graph is defined as a closed, connected sequence of edges, in which no edge appears more than once and no node is visited more than once except to return to the starting node, and which constitutes the boundary of a region of the plane that contains no other edges (Harary, 1969). A mesh can equivalently be defined in terms of its nodes.

[2]This was called to my attention by David West.

of other facts, such as that *AB* is shorter than *CA*? One possibility is that the process is one of exhaustive search over all possible comparisons of certain primitive types (such as line segment length comparisons), with uninteresting observations filtered out as they are generated. Another possibility is that heuristics are used to prune such a search.

5. Relation of the model to other proposals

Johnson-Laird (1983) described a method for making inferences without recourse to a rule-based proof procedure. His method uses what he calls *mental models*. These are finite representations of the content of propositions, and are similar to *models* in the sense used in logic, as introduced by Tarski, except for the finiteness restriction. From propositions are constructed finite sets of individuals that are true interpretations of the propositional content. The result of the construction of such a model is a structure which is a true interpretation of other propositions that could be validly inferred from the original propositions by rule-based deduction. Johnson-Laird presents evidence that humans use such finitary models as a vehicle for inference. The "images" employed in the work presented here are similar in function to Johnson-Laird's "mental models" (rather than what he calls "images"), but employ more elaborate constraints that reflect the structure of two-dimensional space rather than just the logical connectives and quantifiers.

Kosslyn and Shwartz (1977) devised a model of imagery that bears some similarity to our proposal, in that it employed a "surface representation" of images that is consistent with our definition of drawing. Their model and ours differ in several ways, however. The major difference is that we have restricted our consideration to the use of a few primitive object types (lines, etc.), combined into larger types by construction processes. Kosslyn and Shwartz deal with more complex objects, such as representations of chairs and cars, that are defined as specific arrays of pixels that can be scaled and translated. A second difference is our explicit consideration of mechanisms of spatial inference methods, and their role in proposing and disposing of hypotheses. Finally, our model is not motivated by specific psychological facts. Theirs, however, attempted to incorporate many features, such as loss of resolution toward the boundaries, that have not been addressed here.

Funt (1977, 1981) constructed a model that reasons with "images." In that system, two dimensional objects were represented as on/off patterns of "cells" on a "retina." Processes were able to detect instabilities of balance and perform a systematic modification of the images that represented movement of the objects, for example by rotating them about a point. Collisions could be

detected as the result of movements and from these and ancillary assumptions new instabilities could arise. Inferences were then made by playing out a simulation of the movements of the objects represented on the retina. In our terminology, this method is a non-proof-procedural IKR that operates by propagation of constraints. Rather than derive an analytic expression for trajectories, or construct a proof that given initial conditions would lead to certain results, local rules of geometric relations were used to propagate constraints and these constraints then forced the display of conclusions.

Larkin and Simon (1987) describe methods that use "diagrammatic representations" in a problem-solving situation. They illustrate the method with a mechanics problem (involving weights and pulleys) and a geometry problem (involving congruent triangles). They contrast the diagrammatic representation and associated processes with a "sentential representation" of the same problems, showing how the former enjoys decided computational advantages over the latter. The task set for their system is the discovery of a derivation, in the form of an appropriate sequence of production rule applications, of an unknown force ratio, in the mechanics problem case, and of a proof of congruency in the geometry case. The diagrams are used as heuristics in the search for derivations, in the spirit of the Geometry Machine, as discussed above in section 4. In the context of this task, the value of the diagrammatic representation follows from its significant reduction, in comparison to the sentential representation, of the amount of search required to find a derivation. For example, in diagrammatic representations objects are indexed by location and attention moves from one object to adjacent objects; the search order turns out to be substantially better than the essentially exhaustive, unguided order which is all that is offered by the sentential representation. In addition, diagrammatic representations were found to simplify the matching of production rule antecedents to problem features because the diagrams coded the features directly whereas sentential representations hid these features. Since they were requiring the discovery of derivations, rather than simply the "observation" of correct conclusions, Larkin and Simon characterized the major advantages of diagrams as reduction of search and speedup of recognition, stating (p. 71) "... the differential effects on inference appear to be less strong." However, much of what I have been calling inference takes place in their model in the step of producing their "perceptually enhanced data structure," which is a representation that makes "explicit" certain perceptually salient elements, such as alternate interior angles, that are only implicit in the original sentential problem description. Thus their analysis is essentially in agreement with mine and our models have complementary strengths: mine generalizes the methods of diagram construction and retrieval, and theirs interfaces diagrams with rule-based inference.

Related suggestions have been made in the context of knowledge representations that are not explicitly image inspired. I have mentioned the inheritance of properties inference mechanism that is employed in frame-based representation schemes, and acknowledged the relation to our concept of non-proof-procedure inference methods. Elliot (1965), Brown (1970), and Lindsay (1973), employed generalizations of this notion that used properties other than the transitivity of set inclusion. Constraint propagation methods of problem solving have been employed extensively in artificial intelligence, for example, de Kleer (1979) and Steele and Sussman (1978). The truth maintenance system proposal of Doyle (de Kleer, 1984; Doyle, 1979) addresses related issues from a different perspective (maintaining a consistent set of beliefs), but one that is related to our suggested mechanisms for proposing and disposing of conjectures. These and other proposals are compatible with the model outlined in this paper, and offer possible avenues for integrating imagery and more traditional views of knowledge representation in fruitful ways.

There is an extensive literature on machine vision (Brady, 1982; Rosenfeld & Kak, 1976) that deals with systems that attempt to transform digitized physical images into propositional descriptions of "scenes" so that deductions may be made about the *content* of the original images. The output from such "scene analysis" is typically expressed in propositional form: *there is a tank at coordinates (3, 5), this is a view of a ³/₈ inch hex nut from 45° perspective,* and so forth. This is the form of knowledge needed for search and rule-based deduction by artificial intelligence programs that plan (Wilkins, 1984) or problem solve (Nilsson, 1980), for example. Scene analysis, including object recognition, is the process of constructing complex retrieval processes and using them in goal-directed contexts, as discussed in section 4. Ullman (1984) has called such processes *visual routines* and envisions a model of perceptual recognition that constructs complex visual routines from a set of elementary ones. My primitive retrieval processes would presumably comprise a proper subset of the elementary visual routines that Ullman seeks.

6. Discussion

The conventional lay wisdom is that knowledge representations can be divided into two distinct types. The two are imagery and language, corresponding to two phenomenologically distinct objects of introspection. Various suggestions have been put forth in the scientific literature to clarify this fundamental distinction. Often, visual imagery is identified with pictorial rep-

resentations, and language with descriptional representations. Descriptionalists such as Pylyshyn (1980) hold specifically that all knowledge can be represented descriptionally. Apparently no one holds that the pictorial subsumes the descriptional; that battle has been conceded. Thus the question usually addressed is whether visual imagery employs pictorial representations that cannot be fully reduced to descriptions (see Block, 1981, pp. 1–16).

The distinction between pictorial and descriptional representations has proven to be difficult to characterize. One popular approach is to identify pictorial with analog and descriptional with digital (see Kosslyn et al., 1979; also Pylyshyn, 1981). The analog/digital contrast is often in turn taken to be a contrast between continuous and discrete representations. Some writers, myself included, feel that this fails to capture the original imagery/propositional distinction.

Dretske (1981) suggests retaining the "analog–digital" terminology, but glosses it differently: a signal (representational element or notation) carries information that "property *s* has value *F*" in *digital* form if and only if the signal carries no additional information about property *s*; if the signal carries additional information, then by definition it carries information in *analog* form. Under these definitions, knowledge represented propositionally, e.g., as a statement in first-order logic (FOL), represents *only* what is stated explicitly; all other information must be derived by use of a separate set of structural relations among FOL statements, usually in the form of axioms, variable bindings, and rules of inference. On the other hand, a picture of a situation conveys additional information since the representation *must* make some additional things explicit in order to be a picture. To use Dretske's example (p. 137), "The cup has coffee in it" carries no information about how much coffee, how dark it is, or the shape of the cup's handle, whereas a picture of the situation must contain some such additional information. As in my discussion, Dretske's analysis and the intuitions on which it is based emphasize inference as the essential distinction between pictorial and descriptional representations.

A second influential analysis of knowledge representation issues comes to what I take to be the same conclusion. Palmer (1978) proposes a hierarchy of types of "isomorphism" between representation and that which is represented. Physical isomorphisms preserve information by virtue of representing relations that are *identical* to the relations represented. Thus a physical model of a natural terrain preserves the spatial relations of the represented terrain with the very same relations, including for example elevation, but on a smaller scale. Functional isomorphisms, on the other hand, preserve information by representing relations that have the same (algebraic) structure as

the relations represented. Thus the elevations of a natural terrain may be represented as colors on a map of the terrain, provided the colors are interpreted appropriately (as an ordered set) and mapped so as to preserve the order of the physical elevations of the terrain. Thus physical models are a proper subset of functional models. Palmer proposes introducing a class of isomorphisms between physical and functional, which he calls natural isomorphisms. In a natural isomorphism, the representation of preserved relations need not be by means of *identical* relations, as in physical isomorphism, hence not all natural isomorphisms are physical isomorphisms. On the other hand, not just any identically structured set of relations qualify. In a natural isomorphism, the representing constructs have *inherent constraints* (Palmer's term); that is, there is a structure imposed on the representing objects that limits the ways in which they may relate. If these inherent constraints preserve the relations of the represented world, we have a natural isomorphism.

Palmer identifies natural (including physical) isomorphisms with analog (including pictorial) representations, and functional but non-natural isomorphisms with propositional representations. Propositional representations are thus less restricted, as we normally suppose, because the structure of the representing world is extrinsic to it, that is it may be arbitrarily imposed, say in the form of rules of deduction. However, analog (including pictorial) representations employ representations that have inherent (non-arbitrary, unalterable) structure ("inherent constraints") that allow us to do away with deduction rules. This limits their applicability, but at the same time increases their power by reducing the computational complexity of inference (and easing the frame problem). The analysis in this paper, I believe, illustrates how this can be done with a limited class of representation problems, and captures what is common to the Dretske and Palmer definitions.

7. Conclusion: Toward a functional theory of imagery

This paper has proposed a characterization of mental visual imagery that distinguishes that class of knowledge representations from representations based on logical calculi, which are the common currency of artificial intelligence problem-solving programs and expert systems. The proposed knowledge representations may be descriptional and discrete or they may not; they are specified in terms of the mechanism of inference they support. Even in-principle wholly descriptional knowledge representations such as diagrams have important inference support roles that transcend proposition-plus-proof-procedure representations.

The theory defines a set of primitively recognizable features of the class of representation structures, rules for construction that maintain specified values of specified features, and strategies for searching the representation structure for feature values. An example of such a theory has been presented. Each of its components could acquire independent empirical and logical support. The theory does not depend upon knowing whether the substrate of images is neural or electronic.

There are many properties of drawings that are not encompassed by this model; some of them may prove amenable to generalizations of its methods. For example, no use was made of area information. By adding retrieval processes that could compute this metric property, conclusions about equality of area would be possible. Thus certain proofs of the Pythagorean Theorem could be represented and "understood," perhaps even discovered, with such an extension. Textured and colored regions could also be represented by allowing regions to be filled with points of various "colors" or qualities (in addition to "marks") at various settings of the covering resolution. Inferences about intersections of regions would then be available in a non-deductive, constraint generated manner. Inferences about, say, the effects of combining colors would require altogether new sorts of construction and retrieval processes. While I have suggested that the model discussed captures the essential method of the inferential work of imagery, the model presented is limited to very simple, though important, instances – diagrams. Images of more customary experience have not been addressed here. Obviously, extending these notions to percepts/images of complex familiar objects such as faces, animals, and scenes is not straightforward, but the processes dealing with spatial relations here outlined should remain intact and unaffected by the addition of other abilities.

Finally, a word about the status of this work as a psychological model is in order. None of this work has made use of or been tested by experimental methods; it is empirical only in the broad sense of being guided by obvious facts of common sense psychology. It is of relevance to psychological theory, however, in that it addresses a general problem, the frame problem, and an important psychological phenomenon, imagery, for which any theory of natural intelligence must offer some account. One may view perception as offering a solution to the frame problem by allowing "the world" to make appropriate inferences which are then "read" by the brain/mind. If imagery is conceived as a percept-like representation that is evoked in the absence of appropriate sensory input, as the functional equivalence hypothesis of Finke and Shepard has it, then the present model offers an account of how the natural constraints of world situations may be employed to solve the frame problem cognitively.

References

Amarel, S. (1968). On the representation of problems of reasoning about actions. In D. Michie (Ed.), *Machine intelligence 3* (pp. 131–171). New York: American Elsevier.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology.* London: Cambridge University Press.

Block, N. (Ed.). (1981). *Imagery.* Cambridge, Mass.: MIT Press.

Bobrow, D.G. (1975). Dimensions of representation. In D.G. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 1–34). New York: Academic Press.

Bower, G.H. (1972). Mental imagery and associative learning. In L.W. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley.

Brachman, R.J. (1979). On the epistemological status of semantic networks. In N.V. Findler (Ed.), *Associative networks – The representation and use of knowledge in computers.* New York: Academic Press.

Brachman, R.J., & Levesque, H. (Eds.). (1985). *Readings in knowledge representation.* Los Altos, CA: Morgan Kaufmann.

Brady, M. (1982). Computational approaches to image understanding. *Computing Surveys, 14,* 3–71.

Brown, J.S. (1970). *A symbiotic theory formation system.* Unpublished doctoral dissertation, The University of Michigan.

Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81.

de Kleer, J. (1979). *Causal and teleological reasoning in circuit recognition.* (Report No. AI-TR-529). Cambridge, Mass.: Massachusetts Institute of Technology Artificial Intelligence Laboratory.

de Kleer, J. (1984). Choices without backtracking. *Proc. of the National Conference on Artificial Intelligence* (AAAI-84:79–85). Los Altos, Calif.: William Kaufmann.

de Kleer, J., & Brown, J.S. (1984). A qualitative physics based on confluences. In D.G. Bobrow (Ed.), *Qualitative reasoning about physical systems* (pp. 7–83). Amsterdam: Elsevier.

Dennett, D.C. (1981a). The nature of images and the introspective trap. In Block (1981), pp. 51–61.

Dennett, D.C. (1981b). Two approaches to mental imagery. In Block (1981), pp. 87–107.

Doyle, J. (1979). A truth maintenance system. *Artificial Intelligence, 12,* 231–272.

Dretske, F.I. (1981). *Knowledge and the flow of information.* Cambridge, MA: MIT Press.

Elliott, R.W. (1965). A model for a fact retrieval system. Ph.D. dissertation, Austin: The University of Texas.

Evans, G.W. (1980). Environmental cognition. *Psychological Bulletin, 88,* 259–285.

Fahlman, S.E. (1979). *NETL: A system for representing and using real world knowledge.* Cambridge, Mass.: MIT Press.

Fahlman, S.E. (1984). Representing implicit knowledge. In G.E. Hinton & J.A. Anderson (Eds.), *Parallel models of associative memory* (pp. 145–159). Hillsdale, N.J.: Lawrence Erlbaum.

Finke, R.A., & Shepard, R.N. (1986). Visual functions of mental imagery. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance,* Volume II Chapter 37. New York: John Wiley & Sons.

Fodor, J.A. (1981). Imagistic representation. In Block (1981), pp. 63–86.

Forbus, K. (1984). Qualitative process theory. In D.G. Bobrow (Ed.), *Qualitative reasoning about physical systems* (pp. 85–168). Amsterdam: Elsevier.

Funt, B.V. (1977). Whisper: A problem-solving system utilizing diagrams and a parallel processing retina. *Proc. of the 5th International Joint Conference on Artificial Intelligence* (IJCAI-77, 459–464). Pittsburgh: Carnegie-Mellon University.

Funt, B.V. (1981). *Proc. of the 7th International Joint Conference on Artificial Intelligence* (IJCAI-81, 218–220). Menlo Park: AAAI.

Gelernter, H. (1959). Realization of a geometry theorem proving machine. *Proc. International Conference on Information Processing,* pp. 273–282. Paris: UNESCO. Reprinted in E.A. Feigenbaum & J. Feldman (Eds.) (1963). *Computers and thought* (pp. 134–152). New York: McGraw-Hill.

Gelernter, H., Hansen, J.R., & Loveland, D.W. (1960). Empirical explorations of the geometry theorem proving machine. *Proc. Western Joint Computer Conference, 17* (pp. 143–147). Reprinted in E.A. Feigenbaum & J. Feldman (Eds.), (1963), *Computers and thought* (pp. 153–163). New York: McGraw-Hill.

Gibson, J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gladwin, T. (1970). *East is a big bird.* Cambridge, MA: Harvard University Press.

Harary, F. (1969). *Graph Theory.* Reading, Mass.: Addison-Wesley.

Haugeland, J. (1985). *Artificial intelligence. The very idea.* Cambridge, MA: MIT Press.

Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review, 75,* 550–560.

Johnson-Laird, P.N. (1983). *Mental models. Toward a cognitive science of language. inference, and consciousness.* Cambridge, MA: Harvard University Press.

Kosslyn, S.M. (1975). Information representation in visual images. *Cognitive Psychology, 7,* 341–370.

Kosslyn, S.M. (1980). *Images and mind.* Cambridge, Mass.: Harvard University Press.

Kosslyn, S.M., Pinker, S., Smith, G.E., & Shwartz, S.P. (1979). On the demystification of mental imagery. In Block (1981), pp. 131–150.

Kosslyn, S.M., & Pomerantz, J.R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology, 9,* 52–76.

Kosslyn, S.M., & Shwartz, S.P. (1977). A simulation of visual imagery. *Cognitive Science, 1,* 265–295.

Larkin, J., & Simon, H.A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 10,* 65–100.

Lenat, D. (1976). *AM: An artificial intelligence approach to discovery in mathematics as heuristic search.* Stanford, Calif.: Stanford University Computer Science Dept, Report CS-76-57.

Lindsay, R.K. (1961). *Toward the development of machines which comprehend.* Ph.D. Dissertation, Pittsburgh: Carnegie-Mellon University.

Lindsay, R.K. (1963). Inferential memory as the basis of machines which understand natural language. In E.A. Feigenbaum & J. Feldman (Eds.), (1963), *Computers and thought* (pp. 217–233). New York: McGraw-Hill.

Lindsay, R.K. (1973). In defense of ad hoc systems. In R. Schank & K. Colby (Eds.), *Computer models of thought and language* (pp. 372–395). San Francisco: W.H. Freeman.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence,* Volume 4. Edinburgh: Edinburgh University Press.

Nilsson, N. (1980). *Principles of artificial intelligence.* Palo Alto, Calif.: Tioga Press.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, & Winston.

Palmer, S. (1978). Aspects of representation. In E. Rosch & B.B. Lloyd (Eds.), *Computing and categorization* (pp. 259–303). Hillsdale, NJ: Earlbaum.

Posner, M.I., & Keele, S.W. (1967). Decay of visual information from a single letter. *Science, 158,* 137–139.

Pylyshyn, Z.W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *The Behavioral and Brain Sciences, 3,* 111–133.

Pylyshyn, Z.W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88.* In Block (1981), pp. 151–206.

Quillian, R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227–270). Cambridge, Mass.: MIT Press.

Raphael, B. (1971). The frame problem in problem-solving systems. In N.V. Findler & B. Meltzer (Eds.), *Artificial intelligence and heuristic programming.* Edinburgh: Edinburgh University Press.

Reiter, R. (1987). Nonmonotonic reasoning. *Annual Review of Computer Science, 2,* Palo Alto, CA: Annual Reviews, Inc.

Rock, I. (1983). *The logic of perception.* Cambridge, MA: MIT Press.

Rosenfeld, A., & Kak, A.C. (1976). *Digital picture processing*. New York: Academic Press.

Shepard, R.N. (1973). Form, formation, and transformation of internal representations. In R.L. Solso (Ed.), *Contemporary issues in cognitive psychology: the Loyola Symposium*. Washington, D.C.: V.H. Winston & Sons.

Shepard, R.N. (1978). The mental image. *American Psychologist, 33*, 123–137.

Steele, G.L., & Sussman, G.L. (1978). *Constraints*. Cambridge, Mass.: Massachusetts Institute of Technology Artificial Intelligence Laboratory, AIM-502.

Tarjan, R.E. (1971). *An efficient planarity algorithm*. (Report STAN-CS-244-71). Stanford, Calif.: Stanford University.

Ullman, S. (1984). Visual routines. *Cognition, 18*. Reprinted in S. Pinker (Ed.), *Visual cognition* (pp. 97–159). Cambridge, MA: MIT Press.

Wilkins, D.E. (1984). Domain-independent planning: Representation and plan generation. *Artificial Intelligence, 22*, 269–301.

Woodworth, R.S. (1938). *Experimental psychology* (first edition). New York: Holt.

Résumé

On se demande souvent si les images mentales diffèrent fondamentalement des autres formes de représentation des connaissances et, en particulier, des formes prédicatives utilisées dans les programmes d'intelligence artificielle. Parmi les distinctions fréquemment suggérées on trouve: pictural versus descriptif et analogique versus digital. Dans cet article on montre que ces distinctions ne sont cruciales ni pour comprendre le rôle de l'image mentale dans la cognition ni pour cerner la différence entre perception visuelle et langage. On propose de distinguer d'une part la représentation d'image et d'autre part une représentation de la connaissance de type procedure de calcul plus preuve. Cette distinction ne se fonde pas sur des différences de *pouvoir* ou de *forme* d'expression mais sur une distinction entre la *fonction* de ces deux représentations ou plus spécialement sur leur mode d'utilisation pour faire des inférences. Dans cette optique, un rôle fonctionnel important de l'image mentale consiste à fournir une méthode de procédure sans preuve utilisant un mécanisme de satisfaction de contraintes. Les images, même la classe limitée appelée ici diagrammes, confirme l'inférence d'une façon qui diffère de la façon dont le font les représentations prédicatives. Cette analyse offre une approche pour résoudre le "frame-problem" (problème du cadre) en science cognitive.