Problem-Solving with Diagrammatic Representations^{*}

Brian V. Funt

Computer Science Department, State University of New York at Buffalo, Buffalo, NY 14226, U.S.A.**

Recommended by Daniel G. Bobrow and Aaron Sloman

ABSTRACT*

Diagrams are of substantial benefit to WHISPER, a computer problem-solving system, in testing the stability of a "blocks world" structure and predicting the event sequences which occur as that structure collapses. WHISPER's components include a high level reasoner which knows some qualitative aspects of Physics, a simulated parallel processing "retina" to "look at" its diagrams, and a set of re-drawing procedures for modifying these diagrams. Roughly modelled after the human eye, WHISPER's retina can fixate at any diagram location, and its resolution decreases away from its center. Diagrams enable WHISPER to work with objects of arbitrary shape, detect collisions and other motion discontinuities, discover coincidental alignments, and easily update its world model after a state change. A theoretical analysis is made of the role of diagrams interacting with a general deductive mechanism such as WHISPER's high level reasoner.

1. Introduction

Diagrams are very important tools which we use daily in communication, information storage, planning and problem-solving. Their utility is, however, dependent upon the existence of the human eye and its perceptual abilities. Since human perception involves a very sophisticated information processing system, it can be argued that a diagram's usefulness results from its suitability as an input to this powerful visual system. Alternatively, diagrams can be viewed as containing information similar to that contained in the real visual world, the canonical entity the human visual system was presumably designed through evolution to interpret. From this latter perspective, diagrams are a natural representation of certain types of primarily visual information, and the perceptual system simply provides an appropriate set of database accessing functions. Both these viewpoints underly the work described in this paper.

The role of diagrams is explored in a computer problem-solving program, named

* This paper is a substantially lengthened version of a similar paper appearing in IJCAI-5, [6].

Artificial Intelligence 13 (1980), 201–230 Copyright © 1980 by North-Holland Publishing Company

^{**} Now with: Dept. of Computing Science, Simon Fraser University, Vancouver, B.C., Canada V5A IS6.

WHISPER, which refers to diagrams during its processing. WHISPER's high-level reasoning component (HLR). built along the lines of traditional procedural AI problem-solving programs, has the additional option of requesting observations in a diagram. It does this by asking its "perceptual system" to "look at" the diagram with its parallel processing "retina". The questions that the perceptual system can answer are called *perceptual primitives*. If necessary, the HLR can also make changes to the current diagram. Fig. 1 shows WHISPER's overall structure.



FIG. 1. The WHISPER proposal.

Upon receiving a diagram of a blocks world structure, WHISPER outputs a set of diagrams representing the sequence of events which occur as the structure collapses. The HLR contains knowledge about stability and the motion of falling objects. Using the retina to locate objects and their supports, it checks the stability of each object shown in the diagram. Unstable objects may either rotate or slide. In cases where one is rotationally unstable, the HLR asks the retina to "visualize" it rotating and thereby determine at what point it will hit some other object. Using this information WHISPER outputs an updated diagram showing the object rotated into its new position. Then with this new diagram, it restarts the problem-solving process from the beginning—rechecking the stability of each object, moving one of them, outputting another diagram, and restarting again. The process terminates when either all the objects are stable or the problem becomes too complex for the stability tester. A detailed discussion of the HLR will be postponed until Section 3.

1.1. Motivation

A strong case for computer use of diagrams as models for Geometry has been made by Gelernter (1963), and as general analogical representations by Sloman (1971). Networks with nodes representing "ideal integers" and arcs representing relationships between them were used as models for statements in arithmetic by Bundy (1973). Hayes (1974) and Bobrow (1975) comment on the theoretical nature of analogical representations; Hesse (1969) and Nagel (1961) discuss analogical reasoning.

There is a variety of reasons for using diagrams in computer problem-solving. Diagrams such as maps, architectural plans, and circuit diagrams routinely facilitate human problem-solving. Perhaps diagrams function not merely to extend memory capacity, but rather present the important information in a particularly useable form. If they do, then the human visual system provides a paradigmatic example of a system for accessing these representations. Since it exploits a high degree of parallelism, it leads us into the realm of a different type of hardware. This is an exciting step, however, because we can see how much hardware characteristics influence our thinking about the difficulty of various problems and the feasibility of their solution. For example, we know we could compute with Turing Machinesbut would we? Because WHISPER's retina harnesses parallelism, it in effect extends the available machine instruction set with special ones for diagram feature recognition. WHISPER is primarily an exploration of the question: to what extent can problem-solving be simplified through experiment and observation with diagrams? This is in contrast (but not in opposition) to the usual method of deduction within a formal theory as explained in the next section.

1.2. Theoretical framework

Any problem-solving system needs a representation of the problem situation. The standard approach in AI is to formalize the domain. We choose a language and write down a set of statements (axioms, productions, assertions, or a semantic network) describing the world. So that the problem-solver can generate new statements from this initial set, we provide a general deductive mechanism (theorem prover, programming language control structure, network algorithm). In terms of the predicate calculus the axioms define a theory, T, and so long as it is not self-contradictory there will be at least one model M (an assignment of predicates to the predicate symbols, functions to the function symbols, and individuals to the constant symbols) which satisfies it. Since our intention in axiomatizing the world was to accurately describe it, we expect it to be one of the models satisfying T.

We may find a second model M' satisfying T (in general there will be many such models). Now—and this is the main thrust of wHISPER—in some cases we can use M' to provide information about M without deriving it from T. What is required is that some of the predicates, functions and individuals of M' correspond to some of the predicates, functions and individuals of M in such a way that it is possible to translate the results obtained when these predicates and functions are applied to individuals in M' into the results that would be obtained if the corresponding predicates and functions were to be applied to the corresponding individuals in M. The similarity between M and M' means that experiments and observations made in M' yield results similar to those that would be obtained in M. As shown in Fig. 2, for wHISPER M' is the combination of its diagram and diagram re-drawing procedures. wHISPER obtains information about the blocks world by using its retina to observe the results of experimental changes made to its diagrams by the re-drawing procedures.



FIG. 2.

WHISPER is a prototype system designed to explore the extent to which problemsolving can be carried out below the dashed line of Fig. 2; however, it does do some reasoning above the line. WHISPER'S success argues for working below the line, but not against working above the line. WHISPER'S HLR is an above-the-line component.

A natural question is why use M' instead of M? If M is readily accessible then there is no reason not to use it; but, frequently it will not be. For example if we want to determine the stability of a pile of blocks on the surface of the moon, then we could construct a similar pile of blocks on earth and determine the result by experiment. In this case M, the pile of blocks on the moon is inaccessible. We can see that a lot can be learned from the blocks on earth, but some above-the-line inference must be done to handle the discrepancies arising as a result of the difference in gravity.¹

2. Mechanisms for Diagram Interaction

The retina and perceptual primitives are designed to provide WHISPER with a new set of operations whose execution times are of the same order of magnitude as conventional machine instructions. To achieve this a high degree of parallelism has been incorporated into the system. The retina is a parallel processor, and the perceptual primitives are the algorithms it executes. (Do not be misled by the term "retina"; it refers to a general system of receptors and processors for the early stages of perceptual processing, rather than implying any close resemblance to the human retina.) Each perceptual primitive, when executed by the retina, determines whether some particular feature is present in the diagram. WHISPER's retina mixes parallel and sequential computation, so the features it can recognize are not subject to the same theoretical limitations as perceptrons (Minsky and Papert (1969)).

2.1. The retina

WHISPER's retina is a software simulation of hardware which, given the rapidly advancing state of LSI technology, should soon be possible to build. It consists of a collection of processors, each processor having its own input device called a *receptor*. There is a fixed number of processors, and they are all identical. As with the human eye, WHISPER's retina can be shifted to fixate at a new diagram location (also a feature

¹ I am grateful to Raymond Reiter for many of the ideas in Section 1.2.

of a program by Dunlavey (1975)), so that each processor's receptor receives a different input from the diagram. This fixation facility is important because the resolution of the retina decreases from its center to its periphery. Without being able to fixate, it would be impossible for WHISPER to examine the whole diagram in detail. Economy of receptors and processors dictates the use of decreasing resolution. (A declining resolution is also a characteristic of the human eye.) Each receptor covers a separate segment of the diagram and transmits a single value denoting the color of that region. The geometrical arrangement of the receptors and the area each covers is shown in Fig. $3.^2$ The "circles" in the figure are called *bubbles*, and they are



FIG. 3. WHISPER's retina.

arranged in *wedges* (rays emanating from the center) and *rings* (concentric circles of bubbles). The resolution varies across the retina because a larger portion of the underlying diagram is mapped onto a bubble depicted by a larger circle. Since the complete group of receptors is assumed to sense and transmit all signals in parallel, fixations are fast.

Each retinal processor has direct communication links to its nearest neighbors plus one additional link via a common databus connecting all the processors to a supervisory processor called the *retinal supervisor*. The communication topology has been restricted in this simple way to ensure a feasible future hardware implementation.

² There are more receptors filling the central blank area of Fig. 3; however, there is still one special case receptor in the very center which must be handled separately. In order to speed up the retinal simulation the bubbles lying in the blank central area can be fixated separately so they are mapped onto only when they are needed.

The bubble processors are each small computers with independent memory. They all simultaneously execute the same procedure; however, each bubble does not necessarily execute the same instruction at the same time. In the current implementation, a call to the LISP evaluator simulates a processor; and LISP MAPping functions simulate the parallel control structure.

Although the bulk of the processing of the perceptual primitives is done in parallel, there is also a small amount of sequential processing which is performed by the retinal supervisor. The retinal supervisor also directs the parallel processing by choosing which procedure the bubbles should execute next and broadcasting this common procedure to them.

2.2. The perceptual primitives

Each perceptual primitive detects a *problem domain independent* diagram feature. The HLR assigns these features interpretations pertinent to the problem it is solving. The current set of implemented perceptual primitives include ones to: find the center of area of a shape; find the points of contact between a shape of one color and a shape of another; examine curves for abrupt slope changes; test a shape for symmetry; test the similarity of shapes; and visualize the rotation of a shape while watching for a collision with another shape.

The CENTER-OF-AREA perceptual primitive is an illustrative example of the general operation of the perceptual primitives. It computes the center of area of a shape relative to the origin defined by the center of the retina. For each piece, ΔA , of the total area we need to compute the x and y components of its contribution to the total area. Dividing the vector sum of these contributions by the total area yields the coordinates of the center of area. Since each retinal bubble receives its input from a fixed sized area of the diagram and is at a fixed location relative to the retina's center, each bubble can independently compute the components of its contribution to the total area. The bubbles whose receptors do not lie over any part of the shape simply do not contribute. The retinal supervisor performs the summation and the division by the total area. A separate primitive computes the total area. It simply totals the area of all the contributing bubbles. If the computed center of area is far from the retina's center its accuracy can be improved by fixating the retina on the estimated center of area and then recomputing. The decision to iterate is made by the retinal supervisor. The accuracy improves because more of the central, highresolution portion of the retina is used.

It is possible that systematic errors might lead to a discrepancy between the center of area as seen by the retina and the actual center of area of the object in the diagram. This is the case because the diagram-to-retina mapping does not take into account what fraction of a bubble's picture region is covered by an object. The bubble is simply marked whenever any portion of its region is covered. In practice, the accuracy of the center of area test was more than adequate for WHISPER; if necessary the accuracy could always be improved by adding more bubbles to the retina, increasing its resolution. The center of area is used for more than simply providing the center of gravity of the objects in WHISPER's problem domain. Other primitives (symmetry, similarity, and contact finding) fixate on a shape's center of area before beginning their calculations. For example, if a shape is symmetrical its center of area will be on its axis of symmetry.

Another important primitive is RETINAL-VISUALIZATION. What is "visualized" is the rigid rotation of a shape about the retinal center. While the shape is rotating the collision detection primitive can be called as a demon to watch whether the rotation causes the shape to overlap with another stationary shape. This is useful both in "blocks world" environments involving moving objects and in testing whether two shapes are equivalent under rotation. The process is termed *visualization* because it does not involve modifying the diagram, but instead is totally internal to the retina itself. It simply entails an organized and uniform exchange of information amongst neighboring bubbles.

The geometrical arrangement of the bubble receptors facilitates the visualization of rotations. From Fig. 3 it can be seen that aligning the bubble centers along wedges results in a constant angular separation between bubbles of the same ring when they are from neighboring wedges, and that this constant is independent of the ring chosen. Thus, to rotate a shape clockwise each bubble marked by the shape simply sends a message to its clockwise ring neighbor asking it to mark itself. The sender then erases its own mark. A collision is detected if a bubble receives a message to mark when it is already marked by a shape other than the rotating one. Although the shape is rotated in sequential steps, the time required is still short because

(i) there are, as a maximum, only as many steps to be made as there are wedges on the retina (currently 36); and

(ii) all the message passing and collision checking occurs in parallel during each step.

The coarse retinal resolution means that the visualization process is much faster than the alternative of rotating the object by small increments directly in the diagram. However, the coarse resolution also means that the collision test may falsely predict a collision. Although the collision test may occasionally generate such "false alarms", it will never fail to correctly predict a true collision. The reason for this is that during the diagram-to-retina mapping a point in the diagram is blurred to fill a whole bubble on the retina with the result that the objects in the diagram appear slightly enlarged on the retina. To check out a possible false alarm the HLR

(i) calls the re-drawing procedures to rotate the object in the diagram to the point where the collision is expected,

(ii) fixates the retina at the predicted collision point,

(iii) asks the retina (now with its high resolution center) to see if the colliding objects are touching.

The CONTACT-FINDER primitive establishes the points at which an object touches other objects. The retina is first fixated on the center of area of the object and then the retinal supervisor directs each retinal bubble to execute the following steps:

Step 1. If the bubble value is not the color of the object then stop.

Step 2. For each of its neighboring bubbles do Step (3).

Step 3. If neighbor's value is the color of a different object send a "contact-found" message to the retinal supervisor.

Step 4. Stop.

The retinal supervisor may receive quite a number of messages from bubbles in the contact regions. It must sort these into groups—one for each distinct area of contact. To do this the retinal supervisor sequentially follows the chain of neighborhood links from one contact bubble to another. Each bubble in the chain is put in the same contact group. If no neighboring bubble is a contact bubble, then the chain is broken. Long chains indicate that the objects touch along a surface while short ones indicate that they touch only at a point. The bubble coordinates of the endpoints of the chain represent the extremities of a contact surface, and the average of the coordinates of all the bubbles in the group is a good place at which to fixate the retina for a more detailed analysis of the contact.

When two objects touch there is a good chance that one supports the other unless they are just sitting side by side. To determine which object is the supporter and which the supportee, the coordinates of the touching bubbles are compared to find which is "above" the other in the diagram. The assignment of "up" is problem domain dependent and so is made by the HLR.

Another perceptual primitive, FIND-NEAREST, finds the bubble closest to the retinal center satisfying a given condition. For example, to find the object nearest to point P in the diagram the retina is fixated at P and then asked for the nearest marked bubble. The organization of the retina into rings, each an increasing distance from the center, facilitates the search for the required nearest bubble. To find the nearest bubble to the center of the retina satisfying condition C, the retinal supervisor executes the following algorithm:

Step 1. Direct each bubble to test C and save the result (either 'true' or 'false').

Step 2. For n = 1 to the number of rings on the retina do Steps 3 and 4.

Step 3. Direct each bubble to report its wedge and ring coordinates as a message to the retinal supervisor if the following hold: (a) it belongs to ring n, (b) its saved value is 'true'.

Step 4. If there is a message pending for the retinal supervisor from step (3), return the coordinates specified in that message (if there is more than one message pick any one of them—all bubbles in a ring are equidistant from the retinal center) to the calling procedure.

This algorithm is a good example of the difference between efficiency in sequential and parallel computation. Since testing C could be a lengthy computation, it is more efficient in terms of elapsed time to simultaneously test C on all bubbles as in Step 1, than to test it for only those bubbles in the scanned rings of Step 3. On a sequential processor it would be best to test C as few times as possible; whereas, on a parallel processor the total number of times C is tested is irrelevant (assuming the time to compute C(x) is independent of x). It is the number of times C is tested sequentially which is important.

The SYMMETRY primitive tests for symmetry about a designated vertical axis by comparing the values of symmetrically positioned bubbles. An object is symmetrical (WHISPER tests for vertical and horizontal reflective symmetry), if each bubble having its "color" value has a symmetrically located bubble with the same value. If when testing the vertical reflective symmetry of a blue object, say, the bubble in the third wedge clockwise from the vertical axis and in the fourth ring from the center has the value 'blue', then the value of the bubble in the third wedge counterclockwise from the vertical axis and in the fourth ring must be checked to see if it is also 'blue'. If it is not, then possibly the discrepancy can be ruled out as insignificant; otherwise, the object is asymmetrical. Neighborhood message passing is used to bring together the values from bubbles on opposite sides of the proposed axis. The technique is to cause whole wedges to shift in a manner perhaps best described as analogous to the closing of an Oriental hand fan. All the bubbles to the left of the axis send their values clockwise, while all those to the right send theirs counterclockwise. Messages which meet at the axis are compared and will be equal if the object is symmetrical.

The symmetry test must be supplied a proposed axis of symmetry. The center of area offers partial information on determining this axis since it must lie on it if the object is symmetrical. This does not, however, provide the orientation of the axis. Although the simplest solution may be to test the object in all of the wedge orientations by using the rotational visualization, if one more point on the axis of symmetry could be found the axis would be uniquely determined. Such a point is the center of the circumscribing circle of the object. The only problem is that thus far I have not managed to devise a quick parallel algorithm for finding this center. Although in some cases they may coincide, in general I expect the center of area and the center of the circumscribing circle to be distinct for objects with only a single axis of symmetry.

An unexpected and interesting property of WHISPER's retinal geometry leads to a simple method, employing neighborhood communication, for scaling the retinal 'image' of an object. The primitive is RETINAL-SCALING. An object is scaled correctly (i.e. without distorting its shape) if each bubble having its value, sends this value to a bubble in the same wedge, but a fixed number of rings away. As long as each value is moved the same number of rings either inwards or outwards from the bubble which originally holds it, the size of the 'image' of the object is changed but its shape is preserved (Fig. 4). This is the case because the constraint of aligning the bubbles into wedges such that each bubble touches all of its immediate neighbors is satisfied by increasing the bubble diameters by a constant factor from ring to ring. For a proof of this see Funt (1976). Scaling an object by neighborhood communication is implemented by having each bubble simultaneously send its value as a message to its neighbor in the same wedge in either the appropriate inwards or



FIG. 4.

outwards direction, and repeating this message passing process sequentially as many times as necessary to bring about the required scaling.

The SIMILARITY PRIMITIVE determines whether two objects, A and B, are similar under some combination of translation, rotation and scaling, and if so returns the angle of rotation, direction and distance of translation, and the scale factor. It works by taking one object, say A, and translating, scaling and rotating it so it can be matched with the other. Since the center of area of an object is unique the centers of area of A and B must be aligned if they are to match. Thus the first step is to find the centers of area, and then to translate A. Rather than call the re-drawing transformations to move A in the diagram, its translation can be accomplished entirely on the retina by:

(i) fixating on the center of area of A,

(ii) asking all bubbles not containing A to mark themselves as empty space,

(iii) fixating on the center of area of B while superimposing this new image on the old one.

After translation A must be scaled. If A and B are to match, then their areas will need to be the same; therefore, we must scale A by a factor equal to the square root of the ratio of the areas of the two objects (i.e. scalefactor = squareroot(area(B)/area(A)). The areas of A and B are available as a by-product of the center of area calculation. Now that the objects are aligned and the same size, A is rotated about its center of area using retinal visualization to see if there is any orientation at which it matches B.

CURVE-FEATURES analyses curves. In order to begin, it must first find the



Fig. 5.

retinal bubbles on the curve. Given one bubble on the curve, the others can be found by following the chain of bubbles each having the same value. In WHISPER's diagrams the contours of objects are "colored" a different shade from their interiors, and this helps prevent the curve following process from getting lost tracing chains of bubbles which are part of an object's interior. It is not strictly necessary to color code the object contours, since a contour bubble can be determined by the type of neighbors surrounding it, but coding is cheaper and easier.

Once the set of bubbles on the curve is found, each bubble in the set can individually test for the occurrence of a particular feature; therefore, the whole curve is tested in parallel. A bubble detects a sharp bend in the curve if there is an imbalance in the number of its neighbors on opposite sides of the curve which are themselves not members of the curve. This is illustrated by Fig. 5 in which bubble A has three neighbors on each side of the curve, whereas bubble B has six neighbors on one side and none on the other. Thus, a bubble tests for bends by:

(i) asking its neighbors whether or not they are on the curve, and

(ii) comparing the number of responses originating from opposite sides of the curve.

For a simple closed curve, if the bubble knows which responding neighbors are interior and which are exterior, then it can additionally classify the bend as convex or concave.

The slope of a curve at any curve bubble is determined as the perpendicular to the bisector of the angle between the centers of its neighboring bubbles on the curve. This yields a rough approximation to the actual slope, but it is sufficient for quickly testing whether any drastic slope change occurs over the length of the curve. To more accurately determine the slope at a particular point, the retina is fixated on it for higher resolution. The curve tangent is then the perpendicular to the bisector of the angle between wedges with the most bubbles on the curve. The angle between wedges can be used because they emanate directly from the center of the retina, just as the curve must when the retina is centered on it. This method is more accurate than measuring the angle between neighboring bubbles because there are more wedges than neighbors. The HLR mainly uses this test to measure the slope of surfaces at contact points to decide whether or not an object will slide.

2.3. The underlying diagram

We began with the view that the retina is a special purpose parallel processor designed to detect diagrammatic features without saying anything about the precise nature of the diagrams themselves. With the retinal processor in hand, we can now see that the representation of the diagrams is unimportant as long as each bubble receives its correct input. This is analogous to a program which issues a READ command without caring whether the input is coming from a card reader, a file, or a terminal. The method of mapping from the diagram to the retinal bubbles' input must be fast, however, because the retina is re-filled everytime it is fixated at a new diagram location.

There are at least two different types of representing media for the underlying diagram. The first is the conventional medium of visible marks on a two-dimensional surface, usually paper. The map from diagram to human retina is accomplished by the lens of the eye focusing the incoming light. Since there is simultaneous stimulation of the receptors, it is a very fast process.

The second possible type of diagram representation is similar to that used in generating computer graphics. The diagram is specified as a list of primitive elements (in graphics applications, usually line segment equations). In a similar vein, Kosslyn (1975) proposes that human visual imagery is in some ways analogous to the storage and display of graphics images. The parallel processing capacity of WHISPER's retina can be used to quickly map each primitive element into the proper bubble inputs. To mark all bubbles lying on line segment, S, the retinal supervisor directs every bubble to determine independently if it is on S, and if so, to mark itself. Since this simple test—do a circle and a line segment intersect?—is performed by all bubbles simultaneously, the time required is independent of the length of S. The same method can mark all bubbles within any simple shape such as a circle, square or triangle in time independent of its area. Regardless of the type of primitive element, the time taken to "draw" the diagram on the retina is, however, proportional to the number of primitives in its description. They must be processed sequentially.

Due to the lack of true parallel processing, neither of the above two types of diagram representations is used in WHISPER. Instead, the diagram is implemented as a square array. Each array cell denotes a point on a real world, pencil and paper diagram.

2.4. The re-drawing transformations

The re-drawing transformations are the procedures the HLR can call to change the underlying diagram. In WHISPER there are transformations for adding and removing lines, and for rigidly translating and rotating shapes. Other non-linear transformations could be added if required. These re-drawing procedures are of course dependent upon the representation of the diagram they modify, and the ease and efficiency with which they can be implemented could affect the choice of diagram representation.

3. WHISPER in Operation

With the basic mechanisms for interaction with the diagram now understood, it is appropriate to see how they are used in the course of solving a problem. We will consider problems of the type: predict the sequence of events occurring during the collapse of a "blocks world" structure. The structure will be a piled group of *arbitrarily* shaped objects of uniform density and thickness. If the structure is stable, there are no events to describe; if it is unstable, then the events in volve rotations, slides, falls, and collisions. WHISPER accepts a diagram of the initial problem state, and produces a sequence of diagrams, called *snapshots*, as its qualitative solution. A quantitative solution specifying precise locations, velocitites, and times is not found; however, deriving one from a qualitative solution should not be too difficult (deKleer (1975)).

Fig. 6 is a typical example of WHISPER's input diagrams. They all depict a side view of the structure. Each object is shaded a different "color" (alphanumeric value) so it can be easily distinguished and identified. Objects' boundaries are also distinctly colored. The diagram depicts a problem, called the "chain-reaction problem", which is particularly interesting because the causal connection between objects B and D must be discovered.

3.1. The Qualitative HLR

The HLR is the top level of the WHISPER system. It is solely responsible for solving each problem; the diagram and retina are simply tools at its disposal. It consists of procedural specialists which know about stability, about the outcome of different varieties of instability, how t interpret each perceptual primitive, and how to call the transformation procedures to produce the solution snapshots. There are two types of instabilites—rotational and sliding. For clarity, sliding instabilities will not be discussed for the present. Operation of the system follows the steps:

- Step 1. Determine all instabilities.
- Step 2. Pick the dominant instability.
- Step 3. Find pivot point for rotation of unstable object.
- Step 4. Find termination condition of rotation using retinal visualization.
- Step 5. Call transformation procedure to modify diagram as determined in Step 4.



FIG. 6. Chain reaction problem.

Step 6. Output modified diagram as a solution snapshot.

Step 7. Use snapshot from Step 6 as input and restart from Step 1.

In what follows we elaborate on each of these steps.

The diagram and retina are an invaluable aid to the HLR in discovering what stops an object's motion, and in accomplishing the necessary state change. The chain reaction problem demonstrates this. The stability specialist directs the retina to fixate at numerous locations while perusing the diagram, and from an analysis (discussed below) of the visible support relationships determines that B is the only unstable object. B will pivot about the support point closest to its center of gravity. The retina is fixated there (the top right corner of A), so B's rotation can be visualized. As the object rotates, two events are possible. It may collide with another object, or it may begin to fall freely. The conditions under which either of these occur are monitored during the visualization. From this simulation of B's rotation, its collision with D is discovered, and its angle of rotation and location of first contact with D are found. Because of the coarseness of the retinal resolution, this angle of rotation is only approximate. This approximate value is used in conjunction with feedback from the diagram to refine the angle of rotation as follows. First the redrawing transformations are called to produce a new diagram (Fig. 7) in which B



FIG. 7.

	2222222
222222222222222222222222222222222222222	22222222222222 B B22
22222222222 8282 88 8 8 8 8	6 6 6 6 6 6 6 6 8 2 2
2228 8 8 8 8 8 8 8 8 8 8 8	8 8 8 8 8 8 8 8 8 8 2
22228 8 8 8 8 8 8 8 8 8	8 8 8 8 8 8 8 8 8 82
222228 8 8 8 8 8 8 8	B B B B B B B B B B 22
111111111111111228 8 8 8 8 8	8 8 8 8 8 8 8 8 8 8 8
	B B B B B B B B B B B B B B B B B B B
	6 6 6 8 8 8 8 8 22
14 4 4 4 4 4 1 228	8 8 8 8 8 8 8 8 8 8 2
	2 8 8 8 8 8 8 8 8 82
	22222 8 8 8 8 8 2
	22 8 8 88 8 2
	22222B B 2
	222222
	LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL
1A A A A A A A 1	4440 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
	444 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
14 4 4 4 4 4 1	40 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
14 4 4 4 4 4 4 1	444 D D D D D D D D D D D D D D D D 444
14 4 4 4 4 4 1	4440 0 0 0 0 0 0 0 0 0 0 0 0 44
14 4 4 4 4 4 4 1	44D D D D D D D D D D D D A4
14 4 4 4 4 4 1	4444D D D D D D D D D D D D D D40444
1A A A A A A A 1	44446040 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
14 4 4 4 4 4 1	444 D D D D D D D D D D D D D D D D D D
	* ** * ******************************
1A A A A A A A 1	333
1A A A A A A A 1	3 3
14 4 4 4 4 4 1	3 C3
14 4 4 4 4 4 1	3C C33
1A A A A A A A 1	3C C 3
14 4 4 4 4 4 1	33C C C 3
14 4 4 4 4 4 4 1	3 C C C33
	36 6 6 3
14 4 4 4 4 4 1	30 0 0 03
14 4 4 4 4 4 4 1	33C C C C C33
111111111111111	33333333333
999999999999999999999999999999999999	\9999999999999999999999999999999999999

FIG. 8. First snapshot.

is rotated by slightly less than the estimated value. The rotation is made on the short side so that *B* will not overshoot. The retina is then fixated on the anticipated point of collision so that the gap between the two objects can be examined. If there is none, then the update is complete; however, if there is, then *B* is rotated again until the gap is closed. The resulting diagram (Fig. 8) is output as WHISPER's first snapshot of the solution sequence.

3.2. Motion discontinuities and experimental feedback

There are several important observations to be made at this point. One is that discovering the reason for the interruption of an object's motion, accomplished so simply here for *B* through visualization, is generally found to be a very difficult problem. Physics provides equations for object motions, but these equations describe a condition which theoretically lasts indefinitely. They do not indicate when new boundary conditions should take effect. Certainly it is possible to design a set of special heuristics specifying when and where collisions are most likely to occur (e.g. below the rotating object). However, it is quite probable that the collision occurring in Fig. 9 would be overlooked, whereas WHISPER's visualization process would detect it as a matter of course.



Fig. 9.

WHISPER relies on *experimental feedback* to successfully update its diagram in its method of visualization followed by gap closure. This method is basically a pragmatic equivalent to the unfeasible experiment of rotating the object in the diagram by very small increments until a collision occurs. Usually feedback is thought of in terms of a robot immersed in a real world environment. In WHISPER's case, however, the feedback is from a situation analogous to that in the real world—the diagram and diagram transformations—rather than from observation of actual falling objects. Alternatively, we can say that WHISPER is using M' to derive results about M. Using this feedback WHISPER is able to find when and where discontinuous changes in an object's motion occur without being forced to use sophisticated, "numbercrunching" algorithms for touch tests (see Fahlman (1973)) for arbitrary shapes.

3.3. The frame problem

Once WHISPER has produced the first snapshot, it is ready to compute the next one. All the information the HLR needs for this is contained in the first snapshot diagram. Thus to produce the next snapshot, the HLR takes its last output snapshot as input, and begins processing exactly as if it were working on a fresh problem. Although some results derived while working on the previous snapshot remain valid (e.g. some contact relationships will still hold), many will be inapplicable to the new problem. It is easier to disregard this old information than to sort it out and update it, since the retina provides a fast and efficient method of fetching it from the new diagram.

The problem of updating a system's representation of the state of the world to reflect the effects of actions performed in the world is the *frame problem*. Raphael (1971) and Hayes (1971, 1976) discuss it in detail. The transition between WHISPER's snapshots is exactly the type of situation in which we expect the frame problem to arise. It involves the representation of action, the effects of action, and the discovery of chains of causal connection. However, because WHISPER relies on a diagram as a representation of the state of the world, it remains under control. For WHISPER the state of the world is represented by the state of the diagram, and action in the world is represented by corresponding action in the diagram. The corresponding action is the application of the appropriate transformation, and the effects of the action are correctly represented by the resulting state of the diagram.

In WHISPER's current problem, the HLR knows that the action of B's rotation is represented by calling the rotation transformation procedure to re-draw B at its new location in the diagram. Almost all of the information that it needs to continue its problem solving is correctly represented by the updated diagram. It can proceed just as if the new snapshot (the updated diagram) were its original input and it were starting a brand new problem. The most important information which has changed in the transition between the states as a result of the rotation is: the position and orientation of object B; the position of its center of area; the contacts it makes with other objects; and the shape of the areas of empty space. There are also a multitude of things that will not have changed in the world and are correctly left unchanged by the rotational transformation, such as the position of all the other objects, the shape of all objects, the area of all objects, and the contact relationships of other objects not involving B. All of these things work out correctly without the need of any deduction or inference on WHISPER's part. All that it need do is to use its retina to look at the diagram and extract whatever information it needs from the updated diagram.

An expanded WHISPER system could not completely avoid the pitfalls of the frame problem because not all of the information about the current state of the world (e.g. velocities) can be represented by the state of the diagram.

3.4. Subsequent snapshots of the chain reaction problem

The analysis producing the second and third snapshots is very similar to that for the first. In Fig. 8, B's weight on D causes D to be unstable. Its rotation is visualized with the retina fixated at the peak of C leading to the discovery of its collision with the table. The diagram is updated to produce the second snapshot, Fig. 10, which is again input for further analysis. B now lacks sufficient support, and topples to hit D again as shown in Fig. 11. The complexity of the problem rises sharply at this point, and WHISPER's analysis ends, as, I believe, would most peoples'.



FIG. 10. Second snapshot.



FIG. 11. Final snapshot.

B and D could be shown to fall simultaneously (WHISPER currently does not) by rotating D only part of the way to the table before allowing B to catch up, and then iterating this process a few times until D reaches the table.

3.5. Some limitations of WHISPER's qualitative knowledge

WHISPER's knowledge of Finysics is far from comprehensive. As mentioned above, one obvious linitation is that a snapshot by its very nature portrays all objects as stationary, whereas some may be moving. To take velocities into account requires the addition of a quantitative reasoning component to the HLR's qualitative knowledge. Knowledge of velocity, acceleration, momentum and moments of inertia would have to be represented in terms of equations. The HLR's current qualitative predictions can be used to guide the application of these equations in the search for a quantitative solution.

Another limitation is that WHISPER approximates simultaneity by moving objects one after another. This process works for problems like the one discussed above; but this approximation is insufficient in some cases where two or more objects move at a time. In Fig. 12, for example, if B is moved after A is moved, then they will not



FIG. 12.

collide; however, if they are moved simultaneously they will collide. Again we can make use of the diagram by shading the areas each object will sweep through. If no two shaded areas overlap then there will not be a collision; wherever they do overlap a collision might occur and further quantitative analysis of the angular velocities of the objects is required.

3.6. Slide problems

Unstable objects may also slide. When Fig. 13 is given to WHISPER its reasoning up to the point where it generates the first snapshot, Fig. 14, is the same as for the chain reaction problem. At this point it is faced with a problem involving a sliding object. Although the basic outline of the solution process for slide problems—test stability, find termination point of motion, update diagram, output snapshot, restart with the output as input—is the same, there are some essential differences in handling sliding objects. The most important arises because it is not possible to visualize the slide of an object down an arbitrary curve. What WHISPER does instead is examine the curve itself with its retina.

A variety of conditions can terminate an object's slide. For example, there may be a sharp rise (a bump), a sharp fall (a cliff), or a hill which is higher than the starting point. Also the object may slide into another object resting on or near the surface.

11		
11A AL	411	
1A1A A A	A A1A11	
1 4 4 4 4	A A A A11	
	A A A A A11111111111111	11
		A 1111
		* * * * *
		A A A ALAL
		A A A A A AII
<u> </u>		A A A A A All
1		A A A A A A A]
1		* * * * * * * * 1
		A A A A A A A A A
		A A A A A A A A11
11A A A 11 ZHZ		
1 A A A 1 Z B Z	1111	IIA A A A A A A A A
111111111 288	22	1111111111 A A A11
288	82	1111111
28 8 8	82	
26 8 6	5 5 822	
228 8 8	8 5 52	
2888	8882	
28 8 8 8	R R R R 7	
78.8.8.8		<u> </u>
20 8 8 8 8 8		
22 8 8 8 8 8	D D B D D D2	3
		• • • • • • • • • • • • • • • • • • •
2888958		3 C C C C C C C C C C C C C C C C C C C
2888988 28868888	6 6 6 6 6 8 6 82	3 C C C C C C C C C C C 4 D D D D D D D D
2 8 8 8 9 8 8 28 8 8 8 8 8 8 28 8 8 8 8 8 8 8 28 8 8 8	6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	3 C C C C C C C C C C C C C 0 0 0 0 0 0
2 8 8 8 9 8 8 28 8 8 8 8 8 8 28 8 8 8 8 8 8 28 8 8 8	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	3 C C C C C C C C C C C34 D D D D D D D D D D D 0 4 3 C C C C C C C C C C34 D D D D D D D D D D 0 4 3 C C C C C C C C C C34 D D D D D D D D D D 0 4 3 C C C C C C C C C C34 D D D D D D D D D D 0 4 3 C C C C C C C C C C34 D D D D D D D D D D 0 0 0 0 0 0 0 0 0
2 5 5 6 5 5 5 8 26 8 6 5 8 8 8 26 8 6 5 8 8 28 8 6 5 8 8 28 5 8 8 8 8 28 5 8 8 8 8 26 8 8 8 5 8 8	6 6 6 6 6 8 8 82 6 6 6 8 6 6 8 8 82 6 6 6 8 6 6 8 8 8 82 6 8 6 8 6 8 8 6 82 6 8 6 8 6 8 8 8 82	3 C C C C C C C C C C C 34 D D D D D D D D D D D D 0 4 3 C C C C C C C C C C 34 D D D D D D D D D D D 0 4 3 C C C C C C C C C C 34 D D D D D D D D D D 0 4 3 C C C C C C C C C C 34 D D D D D D D D D D D 4 3 C C C C C C C C C C 34 D D D D D D D D D D D 4 3 C C C C C C C C C C 34 D D D D D D D D D D 0 0 0 0 4
2 8 8 8 8 8 8 28 8 8 8 8 8 8 26 8 8 8 8 8 8 26 8 8 8 8 8 28 8 8 8 8 8 28 8 8 8 8 28 8 8 8	6 6 6 6 6 8 6 6 2 6 6 6 6 6 6 8 6 8 2 6 6 6 6 6 6 8 6 8 2 6 6 6 6 6 6 8 6 8 2 6 7 6 6 6 6 8 6 8 8 8 8 8 8 8 8 8 8 8 8	3 C C C C C C C C C C C C 4 0 0 0 0 0 0 0
2 8 8 8 9 8 8 28 8 8 8 8 8 28 8 8 7 9 8 8 28 8 8 7 9 8 28 8 8 8 8 8 28 8 8 9 8 8 2 8 8 9 8 8 2 8 8 9 8 8 2 8 8 7 8 8 8	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	3 C
2 8 8 8 9 8 8 28 8 6 8 8 8 8 28 8 6 8 9 8 28 8 6 9 9 8 28 8 6 9 8 8 2 8 8 9 9 8 2 8 8 9 8 8 2 8 8 7 8 8 8 2 8 6 7 8 8 8	8 <th>$\begin{array}{c} 3 \\ 6 \\ c \\ c$</th>	$\begin{array}{c} 3 \\ 6 \\ c \\ c$
2 8 8 8 8 8 8 8 28 8 6 8 8 8 8 28 8 8 8 7 8 8 28 8 8 7 8 8 28 8 8 8 8 8 28 8 8 9 8 8 2 8 8 7 8 8 2 8 8 7 8 8 2 8 8 7 8 8 8 2 8 8 8 7 8 8 8 2 8 8 8 8 8 8	0 <th>$\begin{array}{c} 3 & C &$</th>	$\begin{array}{c} 3 & C & C & C & C & C & C & C & C & C &$
2 8 8 8 8 8 8 8 28 8 8 5 8 8 8 28 8 8 5 8 8 8 28 8 8 8 8 8 8 28 8 8 8 8 8 2 8 8 8 8	8 8	$\begin{array}{c} 3 \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ \ C \$
2 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	0 0	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
2 8 8 8 8 8 8 8 28 8 6 8 8 8 8 28 8 8 8 7 9 8 8 28 8 8 7 9 8 8 28 8 8 9 9 9 8 8 2 8 8 8 8 8 8 2 8 8 7 8 8 8 2 8 8 7 8 8 8 2 2 8 8 7 8 8 2 2 8 8 7 8 8 2 8 8 8 8 8 2 8 8 8 8 8 2 8 8 8 8 8	0 <th>$\begin{array}{cccccccccccccccccccccccccccccccccccc$</th>	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
2 8 8 8 8 8 8 8 28 8 8 8 8 8 8 28 8 8 8	8 8	$\begin{array}{c} 3 \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ C \ \ \ C \$
2 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	0 0	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
2 8 8 8 8 8 8 8 8 28 8 8 8 8 8 8 8 8 8	0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 2 2 0 0 0 0 0 0 0 2 2 0 0 0 0 0 0 0 0 2 0	$\begin{array}{cccccccccccccccccccccccccccccccccccc$

FIG. 13. A problem with a sliding object.

These conditions are illustrated in Fig. 15. In the current implementation WHISPER fixates only once (at the starting point of the slide) to test for these conditions. Multiple fixations at regular intervals along the curve would improve the accuracy of the tests.

The case shown in Fig. 15(e) of a "surprise collision" is one which requires multiple fixations. Although WHISPER does not yet handle this situation, it is clear how it easily could as illustrated in Fig. 16. In the figure an x indicates a fixation point, a semi-circle indicates the area of the diagram to be checked by the retina at each fixation (checking a circular region is easy because of the retina's ring structure), and the space between the dashed line and the surface indicates a clear "corridor" for the object. The radius of the semi-circle is a function of the object's size and the fixation interval. The same sized corridor can be examined with fewer fixations by using a larger radius. The only disadvantage is that the probability of false alarms is increased because the distance between the dashed line and the circumference of the semi-circles is greater. A false alarm can be investigated by making more fixations in the region where it occurs. This method of detecting collisions is good for two reasons:

(i) because the retina can check large segments of space in a single glance, the number of fixations required to examine the space near the surface is relatively small;

(ii) a collision will never be missed.



FIG. 14.

The HLR, in addition to specifying which conditions the retina should look for, must specify which curve segments it should look at. There are two kinds of curves to test:



FIG. 15.



FIG. 16.

(i) those representing surfaces on the moving object (its underside) which will slide past a point on a stationary object, and

(ii) those representing surfaces of stationary objects (their topsides) which will have a point on the moving object ride over them.

Thus in the current example (Fig. 14) the HLR directs the retina to examine the surface of A from C1 to the left, and the surface of C, possibly continuing over to the surface of D, from C2 to the right. If there is more than one reason why the object's slide will end, then only the condition which occurs first (i.e. after the object has slid the shortest distance) is relevant. One more finepoint is that tests for some conditions, for example collisions, need only be made when the surface is a topside and not an underside.

3.7. Updating the diagram to reflect a slide

After the curve examination is complete and the spot where the object's slide will end is known, the next step is to update the diagram so that it will show the object at its new location. First the HLR calls the re-drawing procedures to translate the object. This is shown in the change from Fig. 14 to Fig. 17 in which point X is aligned with C1. This does not complete the diagram update however, since the object's orientation will most likely change during its slide. The contacts between the object and the surface it slides along should be the same when the slide ends as when it began. Knowing this the HLR can determine the object's correct orientation using retinal visualization. It directs the retina to fixate at the object's new location and then visualize its rotation while watching for the original contact relationships to be re-established. The angle of rotation is returned to the HLR which then calls the re-drawing procedures to rotate the object by that amount in the diagram. As before, the angle returned by the retina is only approximate so the HLR directs the retina to fixate on the expected point of contact and check for any remaining gap. If there is, a second corrective rotation is made. The resulting snapshot is Fig. 18. This two-step method-translation followed by a corrective rotation-works for curved as well as straight surfaces.

What we can see from all this is how experimental feedback combined with a first order theory of sliding motions results in a very natural form of qualitative reasoning.

PROBLEM-SOLVING WITH DIAGRAMMATIC REPRESENTATIONS





3.8. Benefit of the diagram during slide analysis

In the curve examination and diagram update process, the diagram is very useful to the HLR in the course of curve following, and it also provides feedback as it did in the case of rotations. The main pitfall in curve following is the possibility that two objects will coincidentally align so that a smooth curve is formed across them both. An object could then begin its slide on one object and continue sliding along the other as A did when it slid across C and onto D. This *emergent* property of the curve must be noticed, and the two curve segments appended. In a system relying on an independent description of each object, this would pose a significant problem because one would require:

(i) that it have a built in expectation that the situation might arise;

(ii) that it continually check for the situation;

(iii) that its check involve testing whether the first object touches any other object in the universe; and

(iv) that it know how to amalgamate the two separate curve segment descriptions into a new curve description.

For WHISPER it does not create a problem because two aligned surfaces of neighboring objects form a continuous curve in the diagram; WHISPER only has to look at this curve, rather than, in a sense, discover or construct it.

4. The Stability Test

Rather than solve the stability problem with a sophisticated general method, as Fahlman (1973) did in his BUILD system, WHISPER seeks qualitative solutions using rules corresponding to those a person untrained in Physics might apply. The HLR has specialists which express rules like: "If an object hangs over too far, it will topple"; and "If an object and one of its supporters make contact along a surface (rather than at a single point) and if this surface is not horizontal, then the object will slide." A frictionless environment is assumed.

4.1. Sub-structuring

Overall organisation of the stability test is based on the observation that a complete structure is stable if each of its independent subparts is stable whenever their supporters are stable. Thus the initial structure is broken down into smaller substructures whose stability as individual units is easier to test than the stability of the structure as a whole. To perform the stability test the HLR first asks the retina for a list of the names of all objects in the scene. Each object, O, is then handled in turn. The retina is used to find whether or not O supports any other object(s). If it does not, then its stability is tested by SINGLE-STABLE, a specialist in individual object stability. At this point the assumption is that O's supporters are themselves stable. If O supports other objects then its stability is not tested, but rather it is amalgamated with its supportees as if it and they were all glued together to form a single conglomerate object, C. If C does not support anything, then its stability



FIG. 19.

is tested by SINGLE-STABLE; if it does support something, then recursively it is combined with its supportees to form a new conglomerate, C', which is then also checked for supportees.

There is an important exception to the above description. When an object is a cosupporter, as one pillar in an arch for example, then it is not amalgamated with its supportee. In this case the object is sent to SINGLE-STABLE for testing with an addendum specifying the point of contact between it and the supportee.

The dotted curves in Fig. 19 encircle the sub-structures which are passed to SINGLE-STABLE. In (b), Q and RS are cosupporters of X, so they are not combined with it.

Incidentally, treating two objects such as A and B as a single object AB is another example of a situation in which two descriptions must be amalgamated. It is a trivial task for WHISPER to amalgamate two object descriptions, since all it need do is interpret their two color codings as the same color.

4.2. Single object stability

As we have seen, the problem of determining the stability of a complete structure is reduced at each stage to the determination of the stability of individual objects. For a single object there are only three basic types of instability. It can either rotate about some support point (rotational instability), slide along a surface (translational instability), or simply fall freely (free fall instability).

SINGLE-STABLE considers the relative positions of an object's center of gravity and its supporting contacts to decide on rotational stability. Consider first the case of an object with nothing on top of it. One with a single support must have its center of gravity positioned directly above the contact region. One with multiple supports must be positioned so that a vertical dropped from its center of gravity passes through either a contact region or the space between two contact regions. The restrictions of uniform density and thickness of objects mean that an object's center of area can be substituted for its center of gravity. SINGLE-STABLE thus sees that an object "hangs over too far" when its center of area falls outside its supports.



R is stable if $(d_4/d_3)w_R \ge (d_1/(d_1+d_2))w_A$ where: w_A is the weight of A w_R is the weight of R \downarrow is an object's center of gravity

FIG. 20.

The stability of an object with something on top of it will be affected by the extra weight. Because of the way in which objects are formed into conglomerates before they are passed to SINGLE-STABLE, if one supports something then it must in fact be one of two or more cosupporters. Let us say SINGLE-STABLE is testing the stability of an object R which, along with cosupporter S, supports A. First it checks the easy cases:

(i) if ignoring A, R is already rotationally unstable and A's weight will only add to this instability, then R rotates;

(ii) if ignoring A, R is already rotationally unstable, but A's weight might counteract its rotation, then this is a counterbalancing type problem which is too difficult to handle without further quantitative investigation;

(iii) if A, no matter how heavy it is, will not topple R (i.e. test R's stability under the assumption that its center of gravity is located at the contact point between it and A) and R ignoring A is stable, then R remains stable.

The most difficult case is when ignoring A, R by itself is stable, but A may or may not be heavy enough to cause it to rotate. In this situation the location of A's center of gravity relative to its support must be considered. These distances are shown in Fig. 20. If $w_R(d_4/d_3) \ge w_A d_1/(d_1 + d_2)$ then R is stable; otherwise, it will rotate. SINGLE-STABLE cannot handle objects which participate in two or more cosupport relationships. Fig. 21 shows two problems the stability test does not handle.

Objects, such as D in Fig. 6, which are balancing in an unstable equilibrium provide a special problem. Since the slightest deviation in the location of D's center



FIG. 21. Stability problems whisper cannot handle.

of gravity would upset the balance, it must be known precisely. The CENTER-OF-AREA's estimation is not sufficiently accurate. We expect D to balance because it is symmetrical. Using the retina's symmetry test, WHISPER draws the same conclusion. For it to balance, D must be symmetrical about a vertical axis through its support point. Since D is, the stability test reports it as stable; if it were not, then the stability test would have to report that it was unable to decide.

4.3. The eye movement protocol for the chain reaction problem

During the problem-solving process the retina is constantly moving from place to place in the diagram. A trace of the eye movements is given by the circled numbers



FIG. 22.

in Fig. 22 and Fig. 7. Each circle represents a fixation of the retina. The numbers give the order in which the fixations occurred. The retina was split so that its central and peripheral portions could be fixated separately. For some fixations in Fig. 22 only the center of the retina was used, while for others only the periphery. Although moving the two parts of the retina separately would be unnecessary if there actually were many processors operating in parallel, it saves a considerable amount of computation in the pseudo-parallel simulation. What follows is a list of the plotted fixations accompanied by the HLR's reasons for requesting them.

(1) Move to center of diagram; return names of all the objects in the scene.

(2-3) Find the center of gravity of A; find supportees of A.

- (4-5) Find the center of gravity of B; find supportees and supporters of B.
- (6) Move central section of retina; find exact contact point of A and B.

(7-8) Find center of gravity of AB; find supporters of AB.

(9) Move central section; find exact contact point of AB with table.

(10-11) Move central section; find extremities of contact surface.

(12) Find the slope of the contact surface.

(13) Move to center of gravity of B and look at contact between A and B.

(14-15) Move central section; find extremities of contact surface between A and B; (5, 72) and (19, 72) are returned.

(16) Determine the slope of the contact surface.

(17-20) Find center of gravity of D; look for supporters and supportees.

(21-22) Move both the central section and the periphery; find the exact point of contact with C. Discovers that support is a point not a surface indicating possible equilibrium situation.

(23) Move back to center of gravity of D to check for symmetry of D; equilibrium is found to be o.k.

(24) Finding center of gravity of C; look for supportees of C.

(25-26) Finding center of gravity of CD; find supporters of CD; finds the table.

(27) Move central section; find exact point of contact of CD with table.

(28-29) Move central section; find extremities of contact surface; returns (64, 22) and (76, 21).

(30) Determine the type of contact ard its slope.

(31) Move to the pivot point of the rotation of B to visualize the rotation. **** The rotation is then carried out in the diagram, see Fig. 7.****

(32) Move central section to estimated point of collision between B and A to see if they touch; the gap is seen; the amount of the next rotation is estimated. **** Another rotation is carried out in the diagram, see Fig. 8.****

(33) Move central section to estimated point of collision between B and A (the same as (32)); now they are seen to touch.

Although it would be rash to claim that WHISPER accurately models human problemsolving, the eye-movement protocol provides an unusual possibility for testing such a conjecture. An eye-tracking system could be used to record the eye movements of a human subject while he solves one of WHISPER's problems. This record could then be compared with WHISPER's protocol.

4.4 Translational stability

WHISPER decides translational stability by examining the object's contacts. There are three types of contact that are considered: surface-to-surface, surface-to-point, and point-to-surface. The stability criterion for a particular contact is whether or not the tangent to the surface involved in the contact is horizontal at the point of contact. (Tangents are found by the CURVE-FEATURES perceptual primitive.) If the tangent is not horizontal, then the direction of downward tilt is taken as the resultant direction of motion of the object. If a conflict in the direction arises-one contact indicating leftward motion and another indicating rightward motion—then WHISPER reports that it is unable to decide what the motion will be. In these situations a quantitative investigation is needed in order to resolve the qualitative ambiguity. (Resolving qualitative ambiguities by quantitative reasoning is discussed by deKleer (1975).) There is, of course, no conflict between a horizontal contact slope and a non-horizontal contact slope, the former simply does not contribute to the motion. After A rotates to hit C (Fig. 14), the HLR asks the retina to find and classify all the contacts. The A-to-B contact is classified as surface-to-point, with the rightward tilt of the surface of A at the contact noted as contributing to a rightward motion for A. Similarly, the A-to-C contact is classified as point-to-surface with no contribution to the motion of A because the slope of C is horizontal at the contact point. Thus WHISPER concludes that A will slide to the right along the surface of B.

5. Conclusion

WHISPER demonstrates the advantages and feasibility of using diagrams as an aid in problem-solving. We see from a theoretical standpoint that their role is one of a model M' which is similar to the model M—a blocks world structure in the problem domain. More simply stated, WHISPER's diagrams and diagram re-drawing procedures, M', are analogous to blocksworld situations, M. A fundamental component of the system is the retina which blends sequential and parallel processing while limiting the quantity of processors and processor interconnections to a fixed, not too large number. By asking questions of the retina, the HLR is able to obtain experimental feedback from M', and hence results about M—information it would otherwise have to deductively infer from general principles and assertions describing M.

ACKNOWLEDGMENT

I have benefited from the inspiration and assistance of Raymond Reiter, Alan K. Macworth, Richard S. Rosenberg, Gordon McCalla, Peter Rowat, Jim Davidson and Stuart C. Shapiro. I have also received valuable criticism from Zenon Pylyshyn and E. W. Elcock.

REFERENCES

- 1. Bobrow, D., Dimensions of Representation, in: D. G. Bobrow and A. Collins (Eds.), *Representation and Understanding* (Academic Press, New York, 1975), pp. 1-35.
- 2. deKleer, J., Qualitative and Quantitative Knowledge in Classical Mechanics, M.Sc. Thesis (1975), MIT, Cambridge, MA.
- 3. Dunlavey, M., An Hypothesis-Driven Vision System, Advance Papers of the Fourth International Conference on Artificial Intelligence, Tibilisi, Georgia, USSR, September 3-8, 1975, pp. 616-619.
- 4. Fahlman, S., A Planning System for Robot Construction Tasks, AI TR 283, MIT (May 1973).
- 5. Funt, B., WHISPER: A Computer Implementation Using Analogues in Reasoning, Tech. Report 76-9, Department of Computer Science, Univ. of British Columbia (1976).
- 6. Funt, B., WHISPER: A Problem-Solving System Utilizing Diagrams and a Parallel Processing Retina, Advance Papers of the Fifth International Joint Conference on Artificial Intelligence, MIT, August 1977.
- 7. Gelernter, H., Realization of a Geometry-Theorem Proving Machine, in: E. A. Feigenbaum and J. Feldman (Eds.), *Computers and Thought* (McGraw-Hill, New York, 1963), pp. 134–152.
- 8. Hayes, P., A Logic of Actions, in: B. Meltzer and D. Michie (Eds.), *Machine Intelligence 6* (American Elsevier, New York, 1971), pp. 495-520.
- 9. Hayes, P., Some Problems and Non-Problems in Representation Theory. AISB Summer Conference Proceedings (July 1974), pp. 63-79.
- 10. Hesse, Mary, Models and Analogies in Science (University of Notre Dame Press, 1966).
- Kosslyn, S. M., Information Representation in Visual Images, Cognitive Psychology 7 (1975), pp. 341-370.
- 12. Minsky, M. and Papert, S., Perceptrons: An Introduction to Computational Geometry (MIT Press, Cambridge, MA, 1969).
- 13. Nagel, E., The Structure of Science (Harcourt, Brace and World, 1961).
- 14. Sloman, A., Interactions Between Philosophy and Artificial Intelligence: The Role of Intuition and Non-Logical Reasoning in Intelligence, *Artificial Intelligence* 2 (1971), 209-225.
- 15. Raphael, B., The Frame Problem in Problem-Solving Systems, in: N. V. Findler and B. Meltzer (Eds.), Artificial Intelligence and Heuristic Programming (Edinburgh University Press, Edinburgh, 1971), pp. 159-169.

Received 23 January 1979; revised version received 9 July 1979