MINSKY'S FRAME SYSTEM THEORY

Here is the essence of the frame theory: When one encounters a new situation (or makes a substantial change in one's view of a problem), one selects from memory a structure called a frame. This is a remembered framework to be adapted to fit reality by changing details as necessary.

A frame is a data-structure for representing a stereotyped situation like being in a certain kind of living room or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed.

We can think of a frame as a network of nodes and relations. The "top levels" of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals -- "slots" that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet. (The assignments themselves are usually smaller "sub-frames.") Simple conditions are specified by markers that might require a terminal assignment to be a person, an object of sufficient value, or a pointer to a sub-frame of a certain type. More complex conditions can specify relations among the things assigned to several terminals.

Collections of related frames are linked together into frame-systems. The effects of important actions are mirrored by transformations between the frames of a system. These are used to make certain kinds of calculations economical, to represent changes of emphasis and attention, and to account for the effectiveness of "imagery."

For visual scene analysis, the different frames of a system describe the scene from different viewpoints, and the transformations between one frame and another represent the effects of moving from place to place. For non-visual kinds of frames, the differences between the frames of a system can represent actions, cause-effect relations, or changes in conceptual viewpoint. Different frames of a system share the same terminals; this is the critical point that makes it possible to coordinate information gathered from different viewpoints.

Much of the phenomenological power of the theory hinges on the inclusion of expectations and other kinds of presumptions. A frame's terminals are normally already filled with "default" assignments. Thus, a frame may contain a great many details whose supposition is not specifically warranted by the situation. These have many uses in representing general information, most likely cases, techniques for by-passing "logic," and ways to make useful generalizations.

The default assignments are attached loosely to their terminals, so that they can be easily displaced by new items that fit better the current situation. They thus can serve also as "variables" or as special cases for "reasoning by example," or as "textbook cases," and often make the use of logical quantifiers unnecessary.

The frame-systems are linked, in turn, by an information retrieval network. When a proposed frame cannot be made to fit reality -- when we cannot find terminal assignments that suitably match its terminal marker conditions -- this network provides a replacement frame. These inter-frame structures make possible other ways to represent knowledge about facts, analogies, and other information useful in understanding.

Once a frame is proposed to represent a situation, a matching process tries to assign values to each frame's terminals, consistent with the markers at each place. The matching process is partly controlled by information associated with the frame (which includes information about how to deal with surprises) and partly by knowledge about the system's current goals. There are important uses for the information, obtained when a matching process fails; it can be used to select an alternative frame that better suits the situation.

LOCAL AND GLOBAL THEORIES FOR VISION

When we enter a room we seem to see the entire scene at a glance. But seeing is really an extended process. It takes time to fill in details, collect evidence, make conjectures, test, deduce, and interpret in ways that depend on our knowledge, expectations and goals. Wrong first impressions have to be revised. Nevertheless, all this proceeds so quickly and smoothly that it seems to demand a special explanation.

Would parallel processing help? This is a more technical question than it might seem. At the level of detecting elementary visual features, texture elements, stereoscopic and motion-parallax cues, it is obvious that parallel processing might be useful. At the level of grouping features into objects, it is harder to see exactly how to use parallelism, but one can at least conceive of the aggregation of connected "nuclei" (Guzman TR-228), or the application of boundary line constraint semantics (Waltz TR-271), performed in a special parallel network.

At "higher" levels of cognitive processing, however, one suspects fundamental limitations in the usefulness of parallelism. Many "integral" schemes were proposed in the literature on "pattern recognition" for parallel operations on pictorial material -- perceptrons, integral transforms, skeletonizers, and so forth. These mathematically and computationally interesting schemes might quite possibly serve as ingredients of perceptual processing theories. But as ingredients only! Basically, "integral" methods work only on isolated figures in two dimensions. They fail disastrously in coping with complicated, three-dimensional scenery.

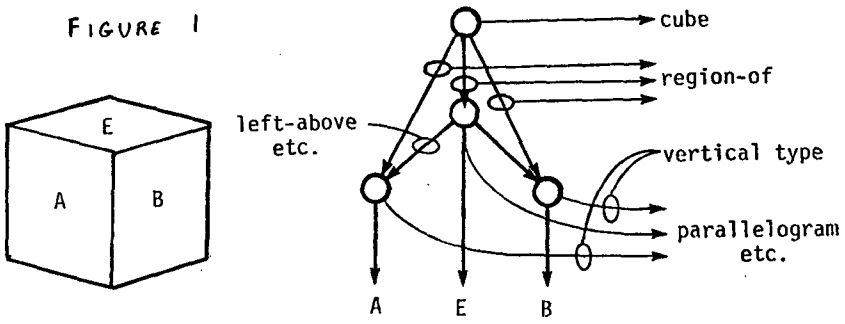
The new, more successful symbolic theories use hypothesis formation and confirmation methods that seem, on the surface at least, more inherently serial. It is hard to solve any very complicated problem without giving essentially full attention, at different times, to different sub-problems. Fortunately, however, beyond the brute idea of doing many things in parallel, one can imagine a more serial process that deals with large, complex, symbolic structures as units! This opens a new theoretical "niche" for performing a rapid selection of large substructures; in this niche our theory hopes to find the secret of speed, both in vision and in ordinary thinking.

SEEING A CUBE

In the tradition of Guzman and Winston, we assume that the result of looking at a cube is a structure something like that in figure 1. The substructures "A" and "B" represent details or decorations on two faces of the cube. When we move to the right, face "A" disappears from view, while the new face decorated with "C" is now seen. If we had to analyse the scene from the start, we would have to

(1) lose the knowledge about "A,"
(2) recompute "B," and
(3) compute the description of "C."

FIGURE 1



But since we know we moved to the right, we can save "B" by assigning it also to the "left face" terminal of a second cube-frame. To save "A" (just in case!) we connect it also to an extra, invisible face-terminal of the new cube-schema as in figure 2.

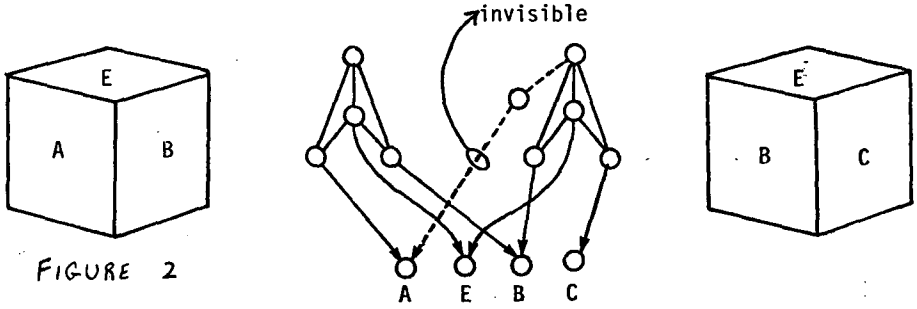


FIGURE 2

If later we move back to the left,

we can reconstruct the first scene without any perceptual computation at all:

just restore the top-level pointers to the first cube-frame. We now need a place to store "C"; we can add yet another invisible face to the right in the first cube-frame! See figure 3. We could extend this to represent further excursions

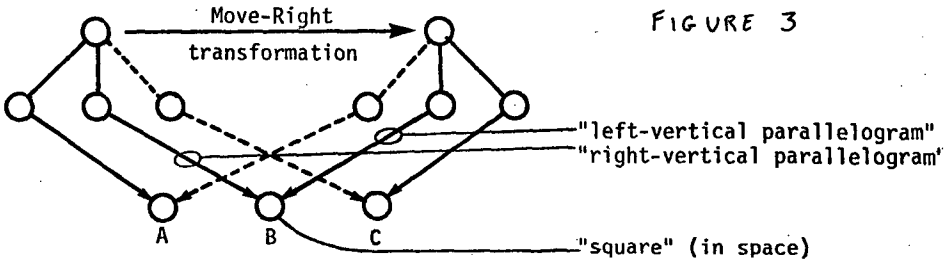


FIGURE 3

around the object. This would lead to a more comprehensive frame system, in which each frame represents a different "perspective" of a cube. In figure 4 there are three frames

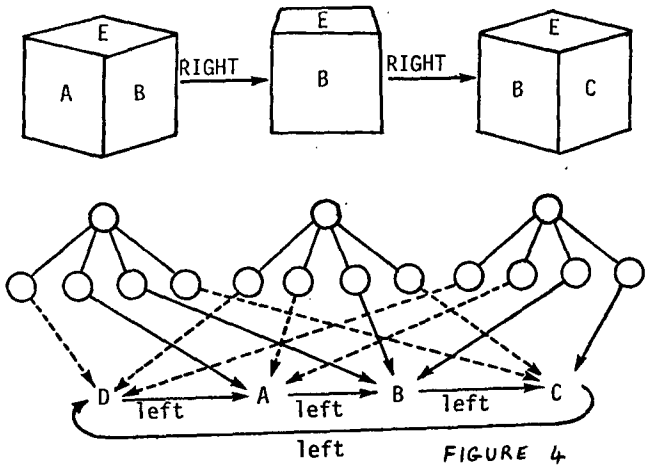


FIGURE 4

corresponding to 45-degree MOVE-RIGHT and MOVE-LEFT actions. If we pursue this analysis, the resulting system can become very large; more complex objects need even more different projections. It is not obvious either that all of them are normally necessary or that just one of each variety is adequate. It all depends.

It is not proposed that this kind of complicated structure is recreated every time one examines an object. It is imagined instead that a great collection of frame systems is stored in permanent memory, and one of them is evoked when evidence and expectation make it plausible that the scene in view will fit it. How are they acquired? We propose that if a chosen frame does not fit well enough, and if no better one is easily found, and if the matter is important enough, then an adaptation of the best one so far discovered will be constructed and remembered for future use.

Each frame has terminals for attaching pointers to substructures. Different frames can share the same terminal, which can thus correspond to the same physical feature as seen in different views. This permits us to represent, in a single place, view-independent information gathered at different times and places. This is important also in non-visual applications.

The matching process which decides whether a proposed frame is suitable is controlled partly by one's current goals and partly by information attached to the frame; the frames carry terminal markers and other constraints, while the goals are used to decide which of these constraints are currently relevant. Generally, the matching process could have these components:

Spatial Frames
.
.
.
.
.
Pictorial Frames
.
.
.
.
.
Relation Markers in common-terminal structure can represent more invariant (e.g. three-dimensional) properties.

(1) A frame, once evoked on the basis of partial evidence or expectation, would first direct a test to confirm its own appropriateness, using knowledge about recently noticed features, loci, relations, and plausible Sub-frames. The current goal list is used to decide which terminals and conditions must be made to match reality.

(2) Next it would request information needed to assign values to those terminals that cannot retain their default assignments. For example, it might request a description of face "C," if this terminal is currently unassigned, but only if it is not marked "invisible." Such assignments must agree with the current markers at the terminal. Thus, face "C" might already have markers for such constraints or expectations as:

* Right-middle visual field.
* Must be assigned.
* Should be visible; if not, consider moving right.
* Should be a cube-face sub-frame.
* Share left vertical boundary terminal with face "B."
* If failure, consider box-lying-on-side frame.
* Same background color as face "B."

(3) Finally, if informed about a transformation (e.g., an impending motion) it would transfer control to the appropriate other frame of that system.

Within the details of the control scheme are opportunities to embed many kinds of knowledge. When a terminal-assigning attempt fails, the resulting error message can be used to propose a second-guess alternative. Later it is shown how memory can be organized into a "Similarity Network" as proposed in Winston's thesis (TR-231).

IS VISION SYMBOLIC?

Can one really believe that a person's appreciation of three-dimensional structure can be so fragmentary and atomic as to be representable in terms of the relations between parts of two-dimensional views? Let us separate, at once, the two issues: is imagery symbolic? and is it based on two-dimensional fragments? The first problem is one of degree; surely everyone would agree that at some level vision is essentially symbolic. The quarrel would be between certain naive conceptions on one side -- in which one accepts seeing either as picture-like or as evoking imaginary solids -- against the confrontation of such experimental results of Piaget (1956) and others in which many limitations that one might fear would result from symbolic representations are shown actually to exist!

As for our second question:

the issue of two- vs. three-dimensions evaporates at the symbolic level.

The very concept of dimension becomes inappropriate. Each type of symbolic representation of an object serves some goals well and others poorly. If we attach the relation labels left-of, right-of, and above between parts of the structure, say, as markers on pairs of terminals, certain manipulations will work out smoothly; for example, some properties of these relations are "invariant" if we rotate the cube while keeping the same face on the table. Most objects have "permanent" tops and bottoms. But if we turn the cube on its side such predictions become harder to make; people have great difficulty keeping track of the faces of a six-colored cube if one makes them roll it around in their mind.

If one uses instead more "intrinsic" relations like next-to and opposite-to, then turning the object on its side disturbs the "image" much less. In Winston's thesis we see how systematic replacements (e.g., of "left" for "behind," and "right" for "in-front-of") can deal with the effect of spatial rotation.

SEEING A ROOM

Visual experience seems continuous. One reason is that we move continuously. A deeper explanation is that our "expectations" usually interact smoothly with our perceptions. Suppose you were to leave a room, close the door, turn to reopen it, and find an entirely different room. You would be shocked. The sense of change would be hardly less striking if the world suddenly changed before your eyes. A naive theory of phenomenological continuity is that we see so quickly that our image changes as fast as does the scene. There is an alternative theory: the changes in one's frame-structure representation proceed at their own pace; the system prefers to make small changes whenever possible; and the illusion of continuity is due to the persistence of assignments to terminals common to the different view-frames. Thus, continuity depends on the confirmation of expectations which in turn depends on rapid access to remembered knowledge about the visual world.

Just before you enter a room, you usually know enough to "expect" a room rather than, say, a landscape. You can usually tell just by the character of the door. And you can often select in advance a frame for the new room. Very often, one expects a certain particular room. Then many assignments are already filled in.

The simplest sort of room-frame candidate is like the inside of a box. Following our cube-model, the room-frame might have the top-level structure shown in figure 5.

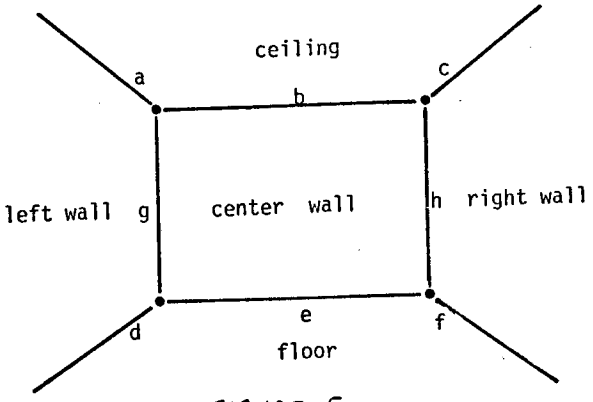


FIGURE 5

One has to assign to the frame's terminals the things that are seen. If the room is familiar, some are already assigned. If no expectations are recorded already, the first priority might be locating the principal geometric landmarks.

To fill in LEFT WALL one might first try to find edges "a" and "d" and then the associated corners "ag" and "gd." Edge "g," for example, is usually easy to find because it should intersect any eye-level horizontal scan from left to right. Eventually, "ag," "gb," and "ba" must not be too inconsistent with one another -- because they are the same physical vertex.

However the process is directed, there are some generally useful knowledge-based tactics. It is probably easier to find edge "e" than any other edge, because if we have just entered a normal rectangular room, then we may expect that

* Edge "e" is a horizontal line.
* It is below eye level.
* It defines a floor-wall texture boundary.

Given an expectation about the size of a room, we can estimate the elevation of "e," and vice versa. In outdoor scenes, "e" is the horizon and on flat ground we can expect to see it at eye-level. If we fail quickly to locate and assign this horizon, we must consider rejecting the proposed frame: either the room is not normal or there is a large obstruction.

The room-analysis strategy might try next to establish some other landmarks. Given "e," we next look for its left and right corners, and then for the verticals rising from them. Once such gross geometrical landmarks are located, we can guess the room's general shape and size. This might lead to selecting a new frame better matched to that shape and size, with additional markers confirming the choice and completing the structure with further details.

SCENE ANALYSIS AND SUBFRAMES

If the new room is unfamiliar, no pre-assembled frame can supply fine details; more scene-analysis is needed. Even so, the complexity of the work can be reduced, given suitable subframes for constructing hypotheses about substructures in the scene. How useful these will be depends both on their inherent adequacy and on the quality of the expectation process that selects which one to use next. One can say a lot even about an unfamiliar room. Most rooms are like boxes, and they can be categorized into types: kitchen, hall, living room, theater, and so on. One knows dozens of kinds of rooms and hundreds of particular rooms; one no doubt has them structured into some sort of similarity network for effective access. This will be discussed later.

A typical room-frame has three or four visible walls, each perhaps of a different "kind." One knows many kinds of walls: walls with windows, shelves, pictures, and fireplaces. Each kind of room has its own kinds of walls. A typical wall might have a 3 x 3 array of region-terminals (left-center-right) x (top-middle-bottom) so that wall-objects can be assigned qualitative locations. One would further want to locate objects relative to geometric inter-relations in order to represent such facts as "Y is a little above the center of the line between X and Z."

In three dimensions, the location of a visual feature of a subframe is ambiguous, given only eye direction. A feature in the middle of the visual field could belong either to a Center Front Wall object or to a High Middle Floor object; these attach to different subframes. The decision could depend on reasoned evidence for support, on more directly visual distance information derived from stereo disparity or motion-parallax, or on plausibility information derived from other frames: a clock would be plausible only on the wall-frame while a person is almost certainly standing on the floor.

Given a box-shaped room, lateral motions induce orderly changes in the quadrilateral shapes of the walls as in figure 6. A picture-frame rectangle, lying flat against a wall, should transform in the same way as does its wall. If a "center-rectangle" is drawn on a left wall it will appear to project out because one makes the default assumption that any such quadrilateral is actually a rectangle hence must lie in a plane that would so project. In figure 7A, both quadrilaterals could "look like" rectangles, but the one to the right does not match the markers for a "left rectangle" subframe (these require, e.g., that the left side be longer than the right side). That rectangle is therefore represented by a center-rectangle frame, and seems to project out as though parallel to the center wall.

Thus we must not simply assign the label "rectangle" to a quadrilateral but to a particular frame of a rectangle-system. When we move, we expect whatever space-transformation is applied to the top-level system will be applied also to its subsystems as suggested in figure 7B.

Similarly the sequence of elliptical projections of a circle contains congruent pairs that are visually ambiguous as shown in figure 8. But because wall objects usually lie flat, we assume that an ellipse on a left wall is a left-ellipse, expect it to transform the same way as the left wall, and are surprised if the prediction is not confirmed.

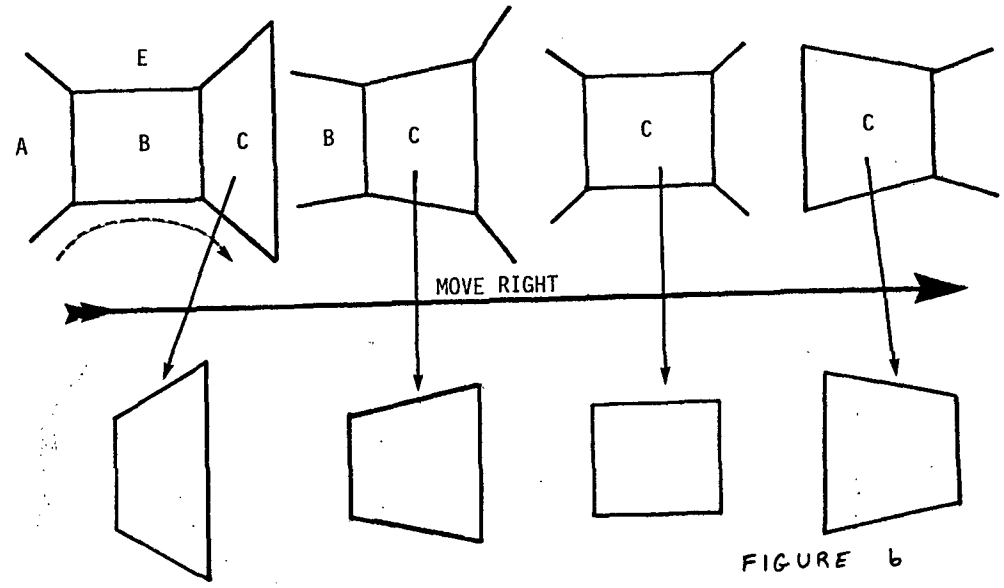
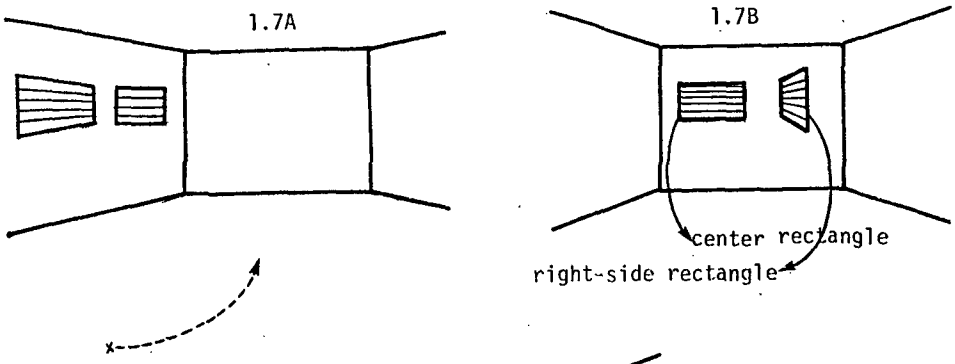


FIGURE 6



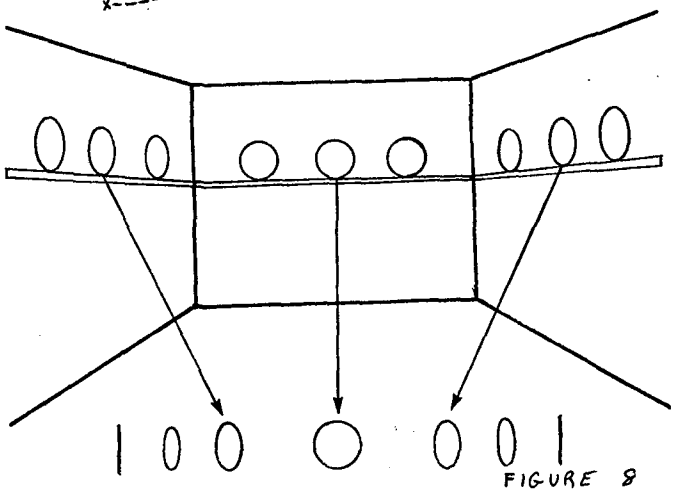


FIGURE 8

DEFAULT ASSIGNMENT

While both Seeing and Imagining result in assignments to frame terminals, Imagination leaves us wider choices of detail and variety of such assignments. Frames are probably never stored in long-term memory with unassigned terminal values. Instead, what really happens is that frames are stored with weakly-bound default assignments at every terminal! These manifest themselves as often-useful but sometimes counter-productive stereotypes.

Thus in the sentence "John kicked the ball," you probably cannot think of a purely abstract ball, but must imagine characteristics of a vaguely particular ball; it probably has a certain default size, default color, default weight. Perhaps it is a descendant of one you first owned or were injured by. Perhaps it resembles your latest one. In any case your image lacks the sharpness of presence because the processes that inspect and operate upon the weakly-bound default features are very likely to change, adapt, or detach them.

WORDS, SENTENCES AND MEANINGS

The concepts of frame and default assignment seem helpful in discussing the phenomenology of "meaning." Chomsky (1957) points out that such a sentence as

(A) "colorless green ideas sleep furiously"

is treated very differently than the non-sentence

(B) "furiously sleep ideas green colorless"

and suggests that because both are "equally nonsensical," what is involved in the recognition of sentences must be quite different from what is involved in the appreciation of meanings.

There is no doubt that there are processes especially concerned with grammar. Since the meaning of an utterance is "encoded" as much in the positional and structural relations between the words as in the word choices themselves, there must be processes concerned with analysing those relations in the course of building the structures that will more directly represent the meaning. What makes the words of (A) more effective and predictable than (B) in producing such a structure -- putting aside the question of whether that structure should be called semantic or syntactic -- is that the word-order relations in (A) exploit the (grammatical) convention and rules people usually use to induce others to make assignments to terminals of structures. This is entirely consistent with grammar theories. A generative grammar would be a summary description of the exterior appearance of those frame rules -- or their associated processes -- while the operators of transformational grammars seem similar enough to some of our frame transformations.

We certainly cannot assume that "logical" meaninglessness has a precise psychological counterpart. Sentence (A) can certainly generate an image! The dominant frame is perhaps that of someone sleeping; the default system assigns a particular bed, and in it lies a mummy-like shape-frame with a translucent green color property. In this frame there is a terminal for the character of the sleep -- restless, perhaps -- and "furiously" seems somewhat inappropriate at that terminal, perhaps because the terminal does not like to accept anything so "intentional" for a sleeper. "Idea" is even more disturbing, because one expects a person, or at least something animate. One senses frustrated procedures trying to resolve these tensions and conflicts more properly, here or there, into the sleeping framework that has been evoked.

Utterance (B) does not get nearly so far because no subframe accepts any substantial fragment. As a result no larger frame finds anything to match its terminals, hence finally, no top level "meaning" or "sentence" frame can organize the utterance as either meaningful or grammatical. By combining this "soft" theory with gradations of assignment tolerances, one could develop systems that degrade properly for sentences with "poor" grammar rather than none; if the smaller fragments -- phrases and sub-clauses -- satisfy subframes well enough, an image adequate for certain kinds of comprehension could be constructed anyway, even though some parts of the top level structure are not entirely satisfied. Thus, we arrive at a qualitative theory of "grammatical:"

if the top levels are satisfied but some lower terminals are not we have a meaningless sentence; if the top is weak but the bottom solid, we can have an ungrammatical but meaningful utterance.

DISCOURSE

Linguistic activity involves larger structures than can be described in terms of sentential grammar, and these larger structures further blur the distinctness of the syntax-semantic dichotomy. Consider the following fable, as told by W. Chafe (1972):

> There was once a Wolf who saw a Lamb drinking at a river and wanted an excuse to eat it. For that purpose, even though he himself was upstream, he accused the Lamb of stirring up the water and keeping him from drinking...

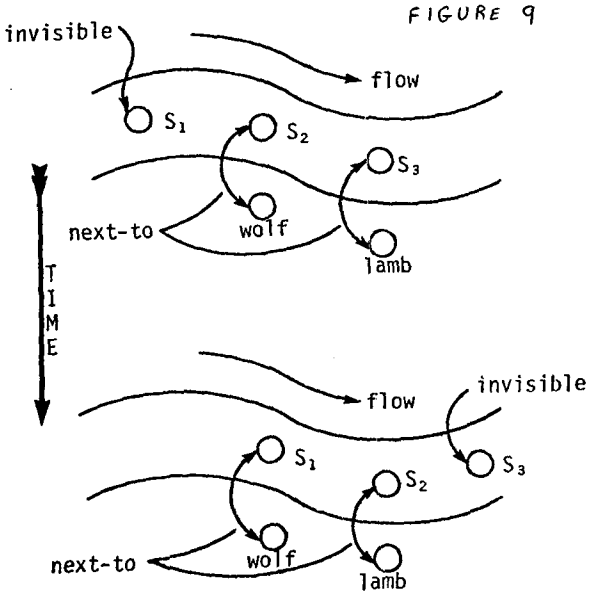
To understand this, one must realize that the Wolf is lying! To understand the key conjunctive "even though" one must realize that contamination never flows upstream. This in turn requires us to understand (among other things) the word "upstream" itself. Within a declarative, predicate-based "logical" system, one might try to formalize "upstream" by some formula like:

```
[A upstream B]
                    AND
[Event T, Stream muddy at A]

                    ===>

[Exists [Event U, Stream muddy at B]]
          AND  [Later U, T]
```

FIGURE 9



But an adequate definition would need a good deal more. What about the fact that the order of things being transported by water currents is not ordinarily changed? A logician might try to deduce this from a suitably intricate set of "local" axioms, together with appropriate "induction" axioms. I propose instead to represent this knowledge in a structure that automatically translocates spatial descriptions from the terminals of one frame to those of another frame of the same system. While this might be considered to be a form of logic, it uses some of the same mechanisms designed for spatial thinking.

In many instances we would handle a change over time, or a cause-effect relation, in the same way as we deal with a change in position. Thus, the concept river-flow could evoke a frame-system structure something like the following, where S1, S2, and S3 are abstract slices of the flowing river shown in figure 9.

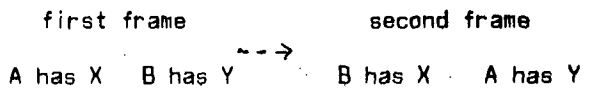
There are many more nuances to fill in. What is "stirring up" and why would it keep the wolf from drinking? One might normally assign default floating objects to the S's, but here S3 interacts with "stirring up" to yield something that "drink" does not find acceptable. Was it "deduced" that stirring river-water means that S3 in the first frame should have "mud" assigned to it; or is this simply the default assignment for stirred water?

Almost any event, action, change, flow of material, or even flow of information can be represented to a first approximation by a two-frame generalized event. The frame-system can have slots for agents, tools, side-effects, preconditions, generalized trajectories, just as in the "trans" verbs of "case grammar" theories, but we have the additional flexibility of representing changes explicitly. To see if one has understood an event or action, one can try to build an appropriate instantiated frame-pair.

However, in representing changes by simple "before-after" frame-pairs, we can expect to pay a price. Pointing to a pair is not the same as describing their differences. This makes it less convenient to do planning or abstract reasoning; there is no explicit place to attach information about the transformation. As a second approximation, we could label pairs of nodes that point to corresponding terminals, obtaining a structure like the "comparison-notes" in Winston (TR-231), or we might place at the top of the frame-system information describing the differences more abstractly. Something of this sort will be needed eventually.

SCENARIOS

We condense and conventionalize, in language and thought, complex situations and sequences into compact words and symbols. Some words can perhaps be "defined" in elegant, simple structures, but only a small part of the meaning of "trade" is captured by:



Trading normally occurs in a social context of law, trust and convention. Unless we also represent these other facts, most trade transactions will be almost meaningless. It is usually essential to know that each party usually wants both things but has to compromise. It is a happy but unusual circumstance in which each trader is glad to get rid of what he has. To represent trading strategies, one could insert the basic maneuvers right into the above frame-pair scenario: In order for A to make B want X more (or want Y less) we expect him to select one of the familiar tactics:

Offer more for Y.
Explain why X is so good.
Create favorable side-effect of B having

Disparage the competition.
Make B think C wants X.

These only scratch the surface. Trades usually occur within a scenario tied together by more than a simple chain of events each linked to the next. No single such scenario will do; when a clue about trading appears it is essential to guess which of the different available scenarios is most likely to be useful.

Charniak's thesis (TR-266) studies questions about transactions that seem easy for people to comprehend yet obviously need rich default structures. We find in elementary school reading books such stories as:

Jane was invited to Jack's Birthday Party.
She wondered if he would like a kite.
She went to her room and shook her piggy bank.
It made no sound.

We first hear that Jane is invited to Jack's Birthday Party. Without the party scenario, or at least an invitation scenario, the second line seems rather mysterious:

She wondered if he would like a kite.

To explain one's rapid comprehension of this, we make a somewhat radical proposal:

to represent explicitly, in the frame for a scenario structure, pointers to a collection of the most serious problems and questions commonly associated with it.

In fact we shall consider the idea that the frame terminals are exactly those questions.

Thus, for the birthday party:

Y must get P for X -------- Choose P!
X must like P ------------ Will X like P?
Buy P ---------- Where to buy P?
Get money to buy P ---- Where to get money?
(Sub-questions of the "present" frame?)
Y must dress up ---------- What should Y wear?

Certainly these are one's first concerns when one is invited to a party.

The reader is free to wonder whether this solution is acceptable. The question "Will X like P?" certainly matches "She wondered if he would like a kite?" and correctly assigns the kite to P. But is our world regular enough that such question sets could be pre-compiled to make this mechanism often work smoothly? The answer is mixed. We do indeed expect many such questions; we surely do not expect all of them. But surely "expertise" consists partly in not having to realize, ab initio, what are the outstanding problems and interactions in situations. Notice, for example, that there is no default assignment for the Present in our party-scenario frame. This mandates attention to that assignment problem and prepares us for a possible thematic concern. In any case, we probably need a more active mechanism for understanding "wondered" which can apply the information currently in the frame to produce an expectation of what Jane will think about.

The key words and ideas of a discourse evoke substantial thematic or scenario structures, drawn from memory with rich default assumptions.

In any event, the individual statements of a discourse lead to temporary representations -- which seem to correspond to what contemporary linguists call "deep structures" -- which are then quickly rearranged or consumed in elaborating the growing scenario representation. In order of "scale," among the ingredients of such a structure there might be these kinds of levels:

EXCUSES

We can think of a frame as describing an "ideal." If an ideal does not match reality because it is "basically" wrong, it must be replaced.

But it is in the nature of ideals that they are really elegant simplifications; their attractiveness derives from their simplicity, but their real power depends upon additional knowledge about interactions between them! Accordingly we need not abandon an ideal because of a failure to instantiate it, provided one can explain the discrepancy in terms of such an interaction. Here are some examples in which such an "excuse" can save a failing match:

OCCLUSION: A table, in a certain view, should have four legs, but a chair might occlude one of them. One can look for things like T-joints and shadows to support such an excuse.

FUNCTIONAL VARIANT: A chair-leg is usually a stick, geometrically; but more important, it is functionally a support. Therefore, a strong center post, with an adequate base plate, should be an acceptable replacement for all the legs. Many objects are multiple purpose and need functional rather than physical descriptions.

BROKEN: A visually missing component could be explained as in fact physically missing, or it could be broken. Reality has a variety of ways to frustrate ideals.

PARASITIC CONTEXTS: An object that is just like a chair, except in size, could be (and probably is) a toy chair. The complaint "too small" could often be so interpreted in contexts with other things too small, children playing, peculiarly large "grain," and so forth.

In most of those examples, the kinds of knowledge to make the repair -- and thus salvage the current frame -- are "general" enough usually to be attached to the thematic context of a superior frame.

ADVICE AND SIMILARITY NETWORKS

In moving about a familiar house, we already know a dependable structure for "information retrieval" of room frames. When we move through Door D, in Room X, we expect to enter Room Y (assuming D is not the Exit). We could represent this as an action transformation of the simplest kind, consisting of pointers between pairs of room frames of a particular house system.

When the house is not familiar, a "logical" strategy might be to move up a level of classification: when you leave one room, you may not know which room you are entering, but you usually know that it is some room. Thus, one can partially evade lack of specific information by dealing with classes -- and one has to use some form of abstraction or generalization to escape the dilemma of Bartlett's commander.

Winston's thesis (TR-231) proposes a way to construct a retrieval system that can represent classes but has additional flexibility. His retrieval pointers can be made to represent goal requirements and action effects as well as class memberships.

> What does it mean to expect a chair? Typically, four legs, some assortment of rungs, a level seat, an upper back. One expects also certain relations between these "parts." The legs must be below the seat, the back above. The legs must be supported by the floor. The seat must be horizontal, the back vertical, and so forth.
> Now suppose that this description does not match; the vision system finds four legs, a level plane, but no back. The "difference" between what we expect and what we see is "too few backs." This suggests not a chair, but a table or a bench.

Winston proposes pointers from each description in memory to other descriptions, with each pointer labelled by a difference marker. Complaints about mismatch are matched to the difference pointers leaving the frame and thus may propose a better candidate frame. Winston calls the resulting structure a Similarity Network.

Is a Similarity Network practical? At first sight, there might seem to be a danger of unconstrained growth of memory. If there are N frames, and K kinds of differences, then there could be as many as K*N*N interframe pointers. One might fear that:

> (1) If N is large, say 10, then N*N is very large -- of the order of 10 -- which might be impractical, at least for human memory.
>
> (2) There might be so many pointers for a given difference and a given frame that the system will not be selective enough to be useful.
>
> (3) K itself might be very large if the system is sensitive to many different kinds of issues.

But, according to contemporary opinions (admittedly, not very conclusive) about the rate of storage into human long-term memory there are probably not enough seconds in a lifetime to cause a saturation problem.

So the real problem, paradoxically, is that there will be too few connections! One cannot expect to have enough time to fill out the network to saturation. Given two frames that should be linked by a difference, we cannot count on that pointer being there; the problem may not have occurred before. However, in the next section we see how to partially escape this problem.

Surface Syntactic Frames --- Mainly verb and noun structures.
Prepositional and word-order indicator conventions.

Surface Semantic Frames ---Action-centered meanings of words.
Qualifiers and relations concerning participants, instruments, trajectories and strategies, goals, consequences and side-effects.

Thematic Frames --- Scenarios concerned with topics, activities, portraits, setting. Outstanding problems and strategies commonly connected with topics.

Narrative Frames --- Skeleton forms for typical stories, explanations, and arguments. Conventions about foci, protagonists, plot forms, development, etc., designed to help a listener construct a new, instantiated Thematic Frame in his own mind.

REQUESTS TO MEMORY

We can now imagine the memory system as driven by two complementary needs.

On one side are items demanding to be properly represented by being embedded into larger frames; on the other side are incompletely-filled frames demanding terminal assignments.

The rest of the system will try to placate these lobbyists, but not so much in accord with "general principles" as in accord with special knowledge and conditions imposed by the currently active goals.

When a frame encounters trouble -- when an important condition cannot be satisfied -- something must be done. We envision the following major kinds of accomodation to trouble.

MATCHING: When nothing more specific is found, we can attempt to use some "basic" associative memory mechanism. This will succeed by itself only in relatively simple situations, but should play a supporting role in the other tactics.

EXCUSE: An apparent misfit can often be excused or explained. A "chair" that meets all other conditions but is much too small could be a "toy."

ADVICE: The frame contains explicit knowledge about what to do about the trouble. Below, we describe an extensive, learned "Similarity Network" in which to embed such knowledge.

SUMMARY: If a frame cannot be completed or replaced, one must give it up. But first one must construct a well-formulated complaint or summary to help whatever process next becomes responsible for reassigning the subframes left in limbo.

MATCHING

When replacing a frame, we do not want to start all over again. How can we remember what was already "seen?" We consider here only the case in which the system has no specific knowledge about what to do and must resort to some "general" strategy. No completely general method can be very good, but if we could find a new frame that shares enough terminals with the old frame, then some of the common assignments can be retained, and we will probably do better than chance.

The problem can be formulated as follows: let E be the cost of losing a certain already assigned terminal and let F be the cost of being unable to assign some other terminal. If E is worse than F, then any new frame should retain the old subframe. Thus, given any sort of priority ordering on the terminals, a typical request for a new frame should include:

(1) Find a frame with as many terminals in common with [a,b,...,z] as possible, where we list high priority terminals already assigned in the old frame.

But the frame being replaced is usually already a subframe of some other frame and must satisfy the markers of its attachment terminal, lest the entire structure be lost. This suggests another form of memory request, looking upward rather than downward:

(2) Find or build a frame that has properties [a,b,...,z]

If we emphasize differences rather than absolute specifications, we can merge (2) and (1):

(3) Find a frame that is like the old frame except for certain differences [a,b,...,z] between them.

One can imagine a parallel-search or hash-coded memory to handle (1) and (2) if the terminals or properties are simple atomic symbols. (There must be some such mechanism, in any case, to support a production-based program or some sort of pattern matcher.) Unfortunately, there are so many ways to do this that it implies no specific design requirements.

Although (1) and (2) are formally special cases of (3), they are different in practice because complicated cases of (3) require knowledge about differences. In fact (3) is too general to be useful as stated, and we will later propose to depend on specific, learned, knowledge about differences between pairs of frames rather than on broad, general principles.

It should be emphasized again that we must not expect magic. For difficult, novel problems a new representation structure will have to be constructed, and this will require application of both general and special knowledge.

CLUSTERS, CLASSES, AND A GEOGRAPHIC ANALOGY

To make the Similarity Network act more "complete," consider the following analogy. In a city, any person should be able to visit any other; but we do not build a special road between each pair of houses; we place a group of houses on a "block." We do not connect roads between each pair of blocks; but have them share streets. We do not connect each town to every other; but construct main routes, connecting the centers of larger groups. Within such an organization, each member has direct links to some other individuals at his own "level," mainly to nearby, highly similar ones; but each individual has also at least a few links to "distinguished" members of higher level groups. The result is that there is usually a rather short sequence between any two individuals, if one can but find it.

At each level, the aggregates usually have distinguished foci or capitols. These serve as elements for clustering at the next level of aggregation. There is no non-stop airplane service between New Haven and San Jose because it is more efficient overall to share the "trunk" route between New York and San Francisco, which are the capitols at that level of aggregation.

The non-random convergences and divergences of the similarity pointers, for each difference d, thus tend to structure our conceptual world around

(1) the aggregation into d-clusters
(2) the selection of d-capitols

Note that it is perfectly all right to have several capitols in a cluster, so that there need be no one attribute common to them all. The "crisscross resemblances" of Wittgenstein are then consequences of the local connections in our similarity network, which are surely adequate to explain how we can feel as though we know what is a chair or a game -- yet cannot always define it in a "logical" way as an element in some class-hierarchy or by any other kind of compact, formal, declarative rule. The apparent coherence of the conceptual aggregates need not reflect explicit definitions, but can emerge from the success-directed sharpening of the difference-describing processes.

The selection of capitols corresponds to selecting stereotypes or typical elements whose default assignments are unusually useful. There are many forms of chairs, for example, and one should choose carefully the chair-description frames that are to be the major capitols of chair-land. These are used for rapid matching and assigning priorities to the various differences. The lower priority features of the cluster center then serve either as default properties of the chair types or, if more realism is required, as dispatch pointers to the local chair villages and towns.

Difference pointers could be "functional" as well as geometric. Thus, after rejecting a first try at "chair" one might try the functional idea of "something one can sit on" to explain an unconventional form. This requires a deeper analysis in terms of forces and strengths. Of course, that analysis would fail to capture toy chairs, or chairs of such ornamental delicacy that their actual use would be unthinkable. These would be better handled by the method of excuses, in which one would bypass the usual geometrical or functional explanations in favor of responding to contexts involving art or play.

ANALOGIES AND ALTERNATIVE DESCRIPTIONS

Suppose your car battery runs down. You believe that there is an electricity shortage and blame the generator.

The generator can be represented as a mechanical system: the rotor has a pulley wheel driven by a belt from the engine. Is the belt tight enough? Is it even there? The output, seen mechanically, is a cable to the battery or whatever. Is it intact? Are the bolts tight? Are the brushes pressing on the commutator?

Seen electrically, the generator is described differently. The rotor is seen as a flux-linking coil, rather than as a rotating device. The brushes and commutator are seen as electrical switches. The output is current along a pair of conductors leading from the brushes through control circuits to the battery.

The differences between the two frames are substantial. The entire mechanical chassis of the car plays the simple role, in the electrical frame, of one of the battery connections. The diagnostician has to use both representations. A failure of current to flow often means that an intended conductor is not acting like one. For this case, the basic transformation between the frames depends on the fact that electrical continuity is in general equivalent to firm mechanical attachment. Therefore, any conduction disparity revealed by electrical measurements should make us look for a corresponding disparity in the mechanical frame. In fact, since "repair" in this universe is synonymous with "mechanical repair," the diagnosis must end in the mechanical frame. Eventually, we might locate a defective mechanical junction and discover a loose connection, corrosion, wear, or whatever.

One cannot expect to have a frame exactly right for any problem or expect always to be able to invent one. But we do have a good deal to work with, and it is important to remember the contribution of one's culture in assessing the complexity of problems people seem to solve. The experienced mechanic need not routinely invent; he already has engine representations in terms of ignition, lubrication, cooling, timing, fuel mixing, transmission, compression, and so forth. Cooling, for example, is already subdivided into fluid circulation, air flow, thermostasis, etc. Most "ordinary" problems are presumably solved by systematic use of the analogies provided by the transformations between pairs of these structures. The huge network of knowledge, acquired from school, books, apprenticeship, or whatever is interlinked by difference and relevancy pointers. No doubt the culture imparts a good deal of this structure by its conventional use of the same words in explanations of different views of a subject.

SUMMARIES: USING FRAMES IN HEURISTIC SEARCH

Over the past decade, it has become widely recognized how important are the details of the representation of a "problem space"; but it was not so well recognized that descriptions can be useful to a program, as well as to the person writing the program. Perhaps progress was actually retarded by ingenious schemes to avoid explicit manipulation of descriptions. Especially in "theorem-proving" and in "game-playing" the dominant paradigm of the past might be schematized so:

> The central goal of a Theory of Problem Solving is to find systematic ways to reduce the extent of the Search through the Problem Space.

Sometimes a simple problem is indeed solved by trying a sequence of "methods" until one is found to work. Some harder problems are solved by a sequence of local improvements, by "hill-climbing" within the problem space. But even when this solves a particular problem, it tells us little about the problem-space; hence yielding no improved future competence. The best-developed technology of Heuristic Search is that of game-playing using tree-pruning, plausible-move generation, and terminal-evaluation methods. But even those systems that use hierarchies of symbolic goals do not improve their understanding or refine their understanding or refine their representations. But there is a more mature and powerful paradigm:

> The primary purpose in problem solving should be better to understand the problem space, to find representations within which the problems are easier to solve. The purpose of search is to get information for this reformulation, not -- as is usually assumed -- to find solutions; once the space is adequately understood, solutions to problems will more easily be found.

The value of an intellectual experiment should be assessed along the dimension of success - partial success - failure, or in terms of "improving the situation" or "reducing a difference." An application of a "method," or a reconfiguration of a representation can be valuable if it leads to a way to improve the strategy of subsequent trials. Earlier formulations of the role of heuristic search strategies did not emphasize these possibilities, although they are implicit in discussions of "planning."

Papert (1972, see also Minsky 1972) is correct in believing that the ability to diagnose and modify one's own procedures is a collection of specific and important "skills." Debugging, a fundamentally important component of intelligence, has its own special techniques and procedures. Every normal person is pretty good at them or otherwise he would not have learned to see and talk! Goldstein (AIM-305) and Sussman (TR-297) have designed systems which build new procedures to satisfy multiple requirements by such elementary but powerful techniques as:

1. Make a crude first attempt by the first order method of simply putting together procedures that separately achieve the individual goals.
2. If something goes wrong, try to characterize one of the defects as a specific (and undesirable) kind of interaction between two procedures.
3. Apply a "debugging technique" that, according to a record in memory, is good at repairing that specific kind of interaction.
4. Summarize the experience, to add to the "debugging techniques library" in memory.

These might seem simple-minded, but if the new problem is not too radically different from the old ones, then they have a good chance to work, especially if one picks out the right first-order approximations. If the new problem is radically different, one should not expect any learning theory to work well. Without a structured cognitive map -- without the "near misses" of Winston, or a cultural supply of good training sequences of problems -- we should not expect radically new paradigms to appear magically whenever we need them.

SOME RELEVANT READING

Abelson, R. P. "The Structure of Belief Systems." Computer Models of Thought and Language. Ed. R. Schank and K. Colby. San Francisco: W. H. Freeman, 1973.

Bartlett, F. C. Remembering. Cambridge: Cambridge University Press, 1967.

Berlin, I. The Hedgehog and the Fox. New York: New American Library, 1957.

Celce-Murcia, M. Paradigms for Sentence Recognition. Los Angeles; Univ. of California, Dept. of Linguistics, 1972.

Chafe, W. First Tech. Report, Contrastive Semantics Project. Berkeley: Univ. of California, Dept. of Linguistics, 1972.

Chomsky, N. "Syntactic Structures." (Originally published as "Strukturen der Syntax") Janua Linguarum Studia Memoriae, 182 (1957).

Fillmore, C. J. "The Case for Case." Universals in Linguistic Theory. Ed. Bach and Harms. Chicago: Holt, Rinehart and Winston, 1968.

Freeman, P. and A. Newell. "A Model for Functional Reasoning in Design." Proc. Second. Intl. Conf. on Artificial Intelligence. London: Sept. 1971.

Gombrich, E. H. Art and Illusion, A Study in the Psychology of Pictorial Representation. Princeton: Princeton University Press, 1969.

Hogarth, W. The Analysis of Beauty. Oxford: Oxford University Press, 1955.

luffman, D. A. "Impossible Objects as Nonsense Sentences." Machine Intelligence 6. Ed. D. Michie and B. Meltzer. Edinburgh: Edinburgh University Press, 1972.

:offka, K. Principles of Gestalt Psychology. New York: Harcourt, Brace and World, 1963.

:uhn, T. The Structure of Scientific Revolutions. 2nd ed. Chicago: University of Chicago Press, 1970.

avoisier, A. Elements of Chemistry. Chicago: Regnery, 1949.

evin, J. A. Network Representation and Rotation of Letters. Publication of the Dept. of Psychology, University of California, La Jolla, 1973.

Minsky, M. "Form and Content in Computer Science." 1970 ACM Turing Lecture. Journal of the ACM, 17, No. 2 (April 1970), 197-215.

Minsky, M. and S. Papert. Perceptrons. Cambridge: M.I.T. Press, 1969.

Moore, J. and A. Newell. "How can MERLIN Understand?" Knowledge and Cognition. Ed. J. Gregg. Potomac, Md.: Lawrence Erlbaum Associates, 1973.

Newell, A. Productions Systems: Models of Control Structures, Visual Information Processing. New York: Academic Press, 1973.

Newell, A. "Artificial Intelligence and the Concept of Mind." Computer Models of Thought and Language. Ed. R. Schank and K. Colby. San Francisco: W. H. Freeman, 1973.

Newell, A. and H. A. Simon. Human Problem Solving. Englewood-Cliffs, N.J.: Prentice-Hall, 1972.

Norman, D. "Memory, Knowledge and the Answering of Questions." Loyola Symposium on Cognitive Psychology, Chicago, 1972.

Papert, S. "Teaching Children to be Mathematicians vs. Teaching about Mathematics." Int. J. Math. Educ. Sci. Technol., 3 (1972), 249-262.

Piaget, J. Six Psychological Studies. Ed. D. Elkind. New York: Vintage, 1968.

Piaget, J. and B. Inhelder. The Child's Conception of Space. New York: The Humanities Press, 1956.

Pylyshyn, Z. W. "What the Mind's Eye Tells the Mind's Brain." Psychological Bulletin. 80 (1973), 1-24.

Roberts, L. G. Machine Perception of Three Dimensional Solids, Optical and Optoelectric Information Processing. Cambridge: M.I.T. Press, 1965.

Sandewall, E. "Representing Natural Language Information in Predicate Calculus." Machine Intelligence 6. Ed. D. Michie and B. Meltzer. Edinburgh: Edinburgh University Press, 1972.

Schank, R. "Conceptual Dependency: A Theory of Natural Language Understanding." Cognitive Psychology (1972), 552-631. see also Schank, R. and K. Colby, Computer Models of Thought and Language. San Francisco: W. H. Freeman, 1973.

Simmons, R. F. "Semantic Networks: Their Computation and Use for Understanding English Sentences." Computer Models of Thought and Language. Ed. R. Schank and K. Colby. San Francisco: W. H. Freeman, 1973.

Underwood, S. A. and C. L. Gates, Visual Learning and Recognition by Computer, TR-123, Publications of Elect. Res. Center, University of Texas, April, 1972.

Wertheimer, M. Productive Thinking. Evanston, Ill.: Harper & Row, 1959.

Wilks, Y. "Preference Semantics." Memo AIM-206, Publications of Stanford Artificial Intelligence Laboratory, Stanford University, July, 1973.

Wilks, Y. "An Artificial Intelligence Approach to Machine Translation." Computer Models of Thought and Language. Ed. R. Schank and K. Colby. San Francisco: W. H. Freeman, 1973.