TWO PROBLEMS IN ANALYZING SCENES

VISION FLASH #12

by

Tim Finin

Massachusetts Institute of Technology

Artificial Intelligence Laboratory

Vision Group

June 1971

ABSTRACT

This paper is based on a B.S. thesis supervised by Patrick Winston. It deals with some previously unexplored problems in the analysis of visual scenes. The scenes consist of two dimensional line drawings of simple objects such as blocks and wedges. The problems have come out of the work that Patrick Winston has done and in discussing them I will be assuming the environment of his system. The first problem asks the questions "When is an object standing? When is it lying?" In the course of answering this question a method is developed for determinig the relative true dimensions of an object from its two dimensional oblique projection. The second problem develops methods for discovering when one object is in front of another in situations where previous methods have failed.

1. THE ATTITUDE OF AN OBJECT

1.0

In Patrick Winston's Ph.D. thesis, <u>Learning Structural Descriptions from Examples</u>, one of the properties that could be assigned to an object was its attitude -- the object could either be standing or lying. However, as of yet, no method has been devised to determine the attitude of an object from its two-dimensional line drawing representation. This problem breaks easily into two parts, namely:

1. Giving the machine a definition of standing and lying that will agree with human judgment.
2. Determining the true dimensions of an object from its two dimensional representation.

Both of these problems turn out to be fairly complex for a general object and even a class of "simple" objects. Consequently, I will address myself first to answering these problems for the "simplest" of solids -- the BRICK (rectangular parallelepiped) -- and then speculate on extensions to other 'simple objects'.

1.1 What Do We Mean by Standing and Lying

Let's assume that we are looking at bricks that are resting on a flat surface. The intuitive definitions of standing and lying seem to be:

A brick is STANDING if it is resting
on one of its smallest sides.

A brick is LYING if it is resting on
one of its largest sides.

So far this is fine, but what do we say if a brick rests on a
side which is neither its largest or smallest?   See figure 1.1.1.
Sometimes we would want to say that such a brick is standing and sometimes
lying, depending on the areas of the other sides, as in figure 1.1.2.

Let's imagine we have long, think boards which we will nail together
to form thicker boards, as in figure 1.1.3.   I would judge that when
we have just one board, it is standing.   With a second board added, I
would still adjuge the resulting object to be standing.   If we increase
the thickness any more, the object seems to be lying.   Let's define the
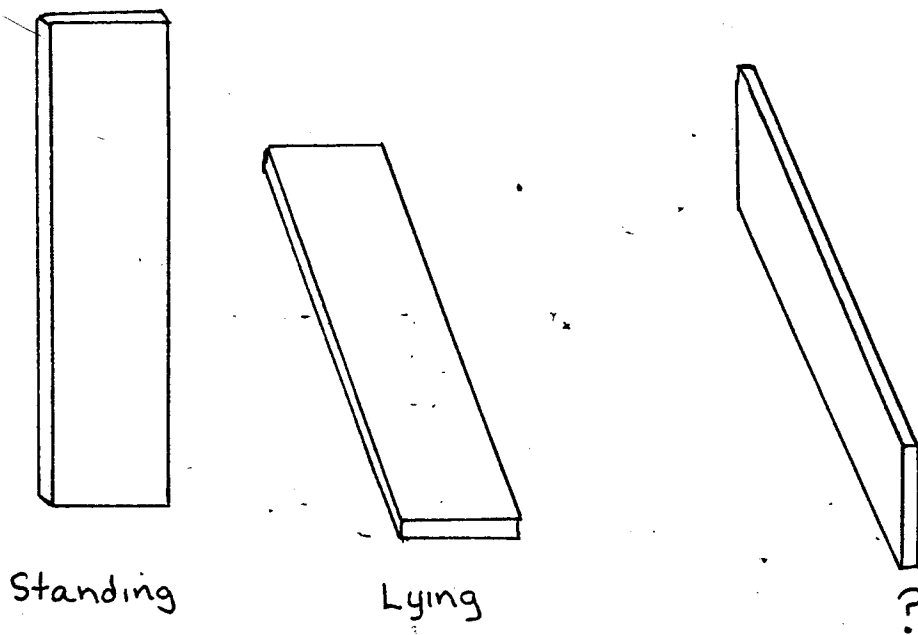ratio R_1 to be equal to the ratio of the largest area to the next-to-largest
area.

Standing   Lying   ?

FIG. 1.1.1

Lying   Standing
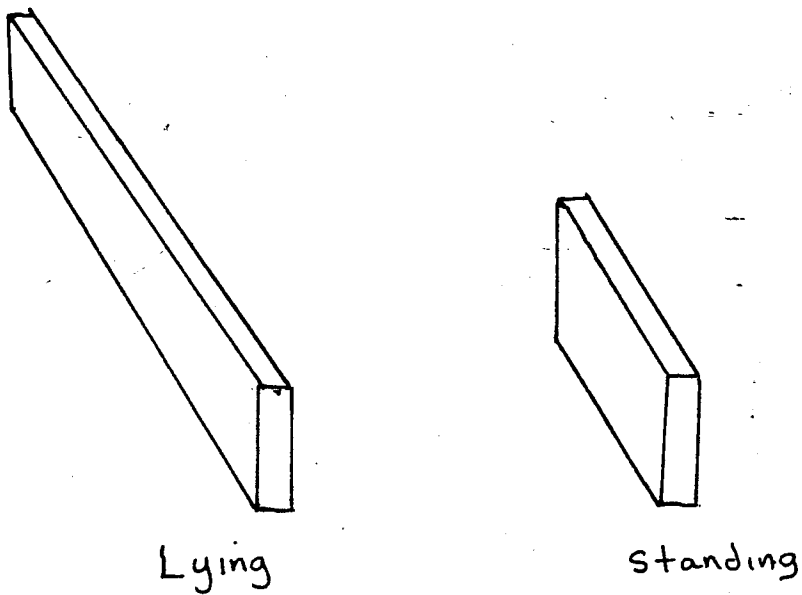
FIG. 1.1.2

$$R_1 = \frac{zy}{zx} = \frac{y}{x}$$

where z = longest side

    y = next-to-longest side

    x = shortest side.

Using this measure, it would seem that when R_1 is greater than about 3, we would say that the object is standing, and when less than three, lying.

This ratio does not take into account the length of our boards however, to let's look at another example. Imagine we have a short board (not too thin) which we can stretch, as in figure 1.1.4. We will define a ratio R_2 to be the ratio of the next-to-largest area to the smallest area.

$$R_2 = \frac{yz}{yx} = \frac{z}{y}$$

It would seem that if R_2 is less than about three, then the brick is standing, and if greater than than three, then the brick is lying.

When a brick of general dimensions is resting on its next-to-largest side we must use both these measures to determine whether it is standing or lying.

We have: $R_1 > 3 \rightarrow$ standing

             $R_1 < 3 \rightarrow$ lying

             $R_2 > 3 \rightarrow$ lying

             $R_2 < 3 \rightarrow$ standing
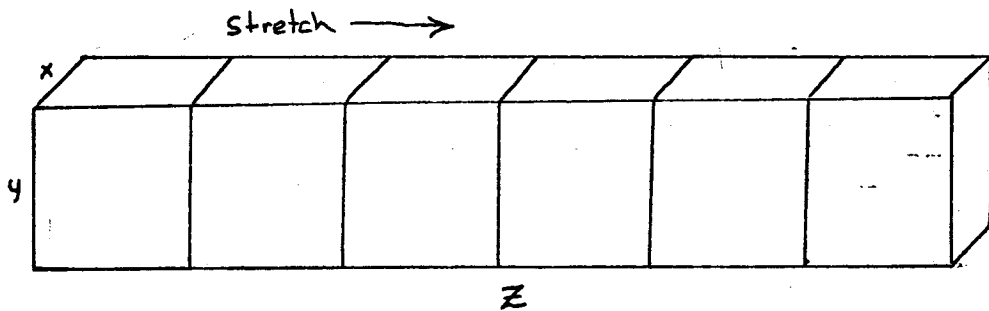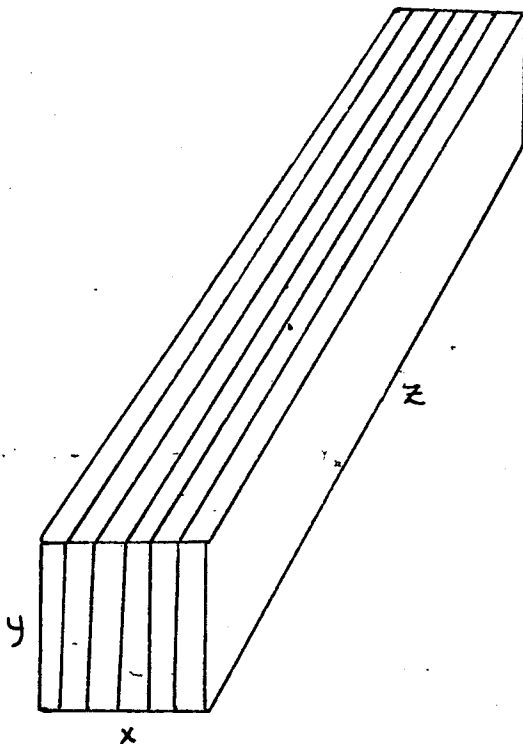
FIG.1.1.3

FIG. 1.1.4

If both our measures agree for an object, fine.   If they contradict
one another, then we will accept whichever evidence is stronger.   The
more extreme the measure, the stronger the evidence.

If: $R_1 > 3$ and $R_2 > 3$ and $R_1 > R_2$ $\longrightarrow$ standing

$R_1 > 3$ and $R_2 > 3$ and $R_1 < R_2$ $\longrightarrow$ lying

$R_1 < 3$ and $R_2 < 3$ and $R_1 > R_2$ $\longrightarrow$ standing

$R_1 < 3$ and $R_2 < 3$ and $R_1 < R_2$ $\longrightarrow$ lying

We can simplify this to one test through a new ratio:

$$R = \frac{R_1}{R_2} = \frac{y^2}{zx} > 1 \longrightarrow \text{standing}$$

$$\frac{y^2}{zx} < 1 \longrightarrow \text{lying}$$

In fact, we can ue this same ratio to define the attitude of
a brick resting on any of its sides.   Let x and z be the dimensions of
the side on which the brick rests and y the other dimension.   The ratio
is then the brick's height squared over the area of the side on which it
rests.

If the object rests on its largest side then:

$$zx > xy \longrightarrow z > y$$

$$zx > yz \longrightarrow x > y$$

so that $R = \frac{y^2}{xz}$ will certaily be less than one, implying that our object
is lying.   If the object rests on its smallest side then:

$$xz < xy \rightarrow z < y$$

$$xz < zy \rightarrow x < y$$

so that $R = \dfrac{y^2}{xz}$ will be greater than one and our object must be standing.

Thus, if we can determine this ratio for a brick, then we decide whether it is standing or lying. Let's see how this can be done.

Using a procedure Winston has developed, we can pick out the lines which form the bottom border of the object. See figure 1.1.5. These lines determine the side on which the brick rests, as well as the denominator of our test ratio. To get the numerator of our ratio we must simply square the other dimension. Any of the lines intersecting one of the bottom lines will suffice to get y^2.

Of course, when R is close to one, its judgment becomes somewhat arbitrary. In these cases it would perhaps be best to say that the object could be either standing or lying. See figuee 1.1.6. In such cases our human judgment allows for both statements of the object's attitude.

This same ratio seems to work well in determining the attitude of objects such as parallelepipieds, pyramids and wedges. For other objects, the problem is more complex. Consider the cylindrical solid in figure 1.1.7. If this object rests on one of its ends, as in figure 1.1.7A, then the ratio of its height squared to the area of its base seems to give a good definition of its attitude.
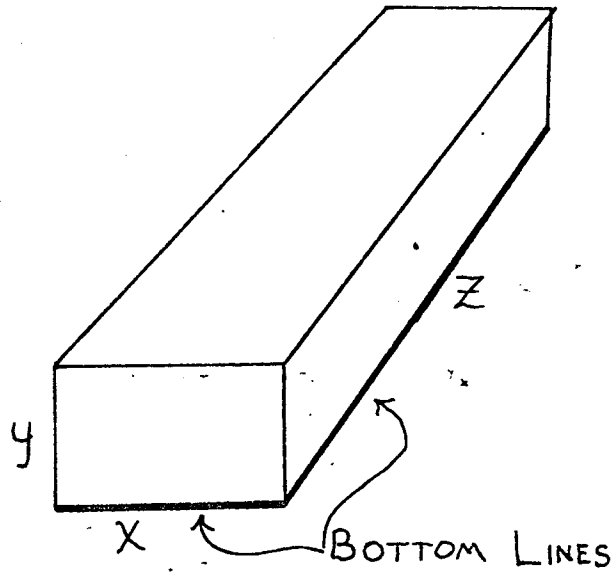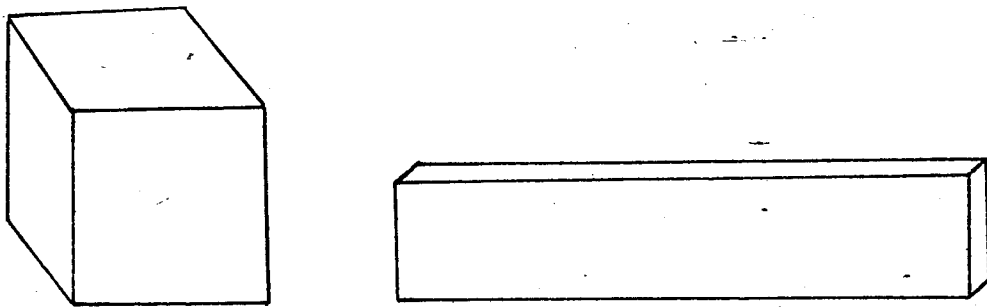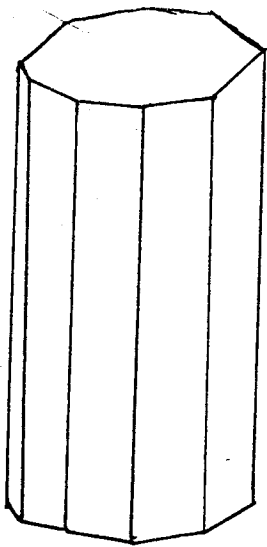
FIG. 1.1.5

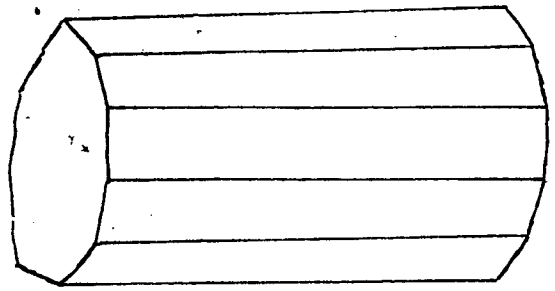$R \approx 1 \Rightarrow$ STANDING OR LYING

FIG. 1.1.6

In figure 1.1.7B, however, the ratio still claims the object is standing, by virtue of the small area on which it rests. I suspect that we might be able to win with such objects if we consider a box just large enough to hold the object. If the box is standing, then the object is standing. If the box is lying, then the object is lying.

Other objects give us even more trouble, such as cones and truncated cones. The nature of these objects exhibits a strong influence on our judgment of standing and lying. Consider the conic section in figure 1.1.8B. I am hesitant to say that it is standing even though it's height is much larger than any of its other dimensions. It would seem that such an object is standing only if it rests on its base and the ratio of its height squared to the area of its base is greater than one (figure 1.1.8A). It would be lying if it did not rest on its base (figures 1.1.8B and 1.1.8C). When such an object rests on its base and the ratio is less than one we would probably be better off not assigning a standing-lying attitude at all (figure 1.1.8D).
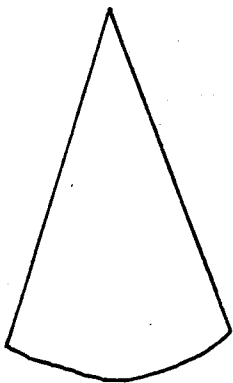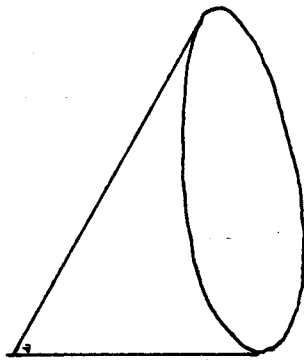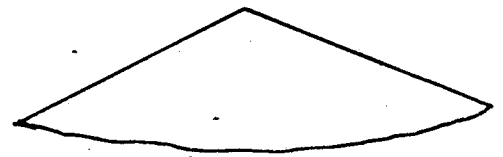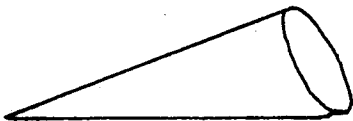
A

B

FIG. 1.1.7

A

B

C

D

FIG. 1.1.8

## 1.2 True Dimensions of a Block from its Oblique Projection

In order to determine whether a block is standing or lying we must know the ratios of its dimensions. In the two dimensional projections we are working with, some, if not all, of the dimensions will appear foreshortened. The amount of foreshortening will depend, of course, on the angle from which we are viewing.

Let's assume we are viewing blocks which rest on a level table. A block in a general oblique position will then appear as in figure 1.2.1. X', Y', and Z' are the apparent dimensions as measured in our projection, and Θ'_x and Θ'_y are the apparent angles X and Y make with a horizontal line. From the angles Θ'_x and Θ'_y we can determine our viewing angle, defined by an angle of elevation, S, and an angle of rotation, R, and thus the amount of foreshortening in our projected drawing. Figure 1.2.2 shows our scene from some different views, the arrow representing our line of sight.
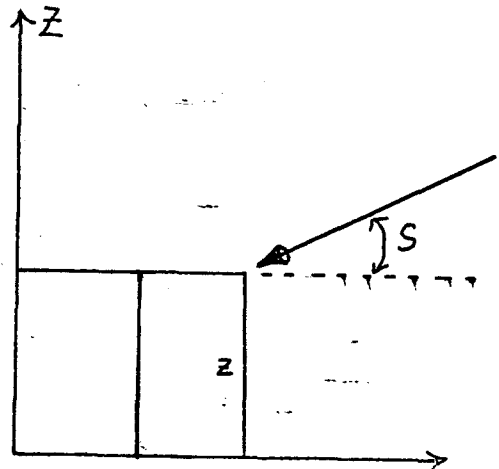
The amount of distortion between the apparent angles Θ'_x and Θ'_y and the true angles Θ_x and Θ_y depends on the elevation angle of our viewpoint. Going through the geometry of our situation we find that:

$$\tan \Theta'_x = \tan \Theta_x \sin S \quad (1)$$

$$\tan \Theta'_y = \tan \Theta_y \sin S \quad (2)$$
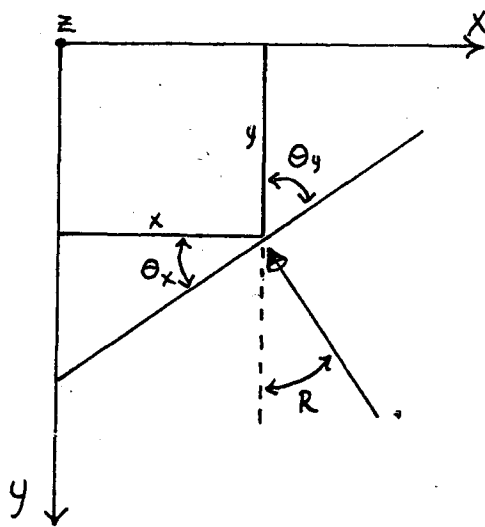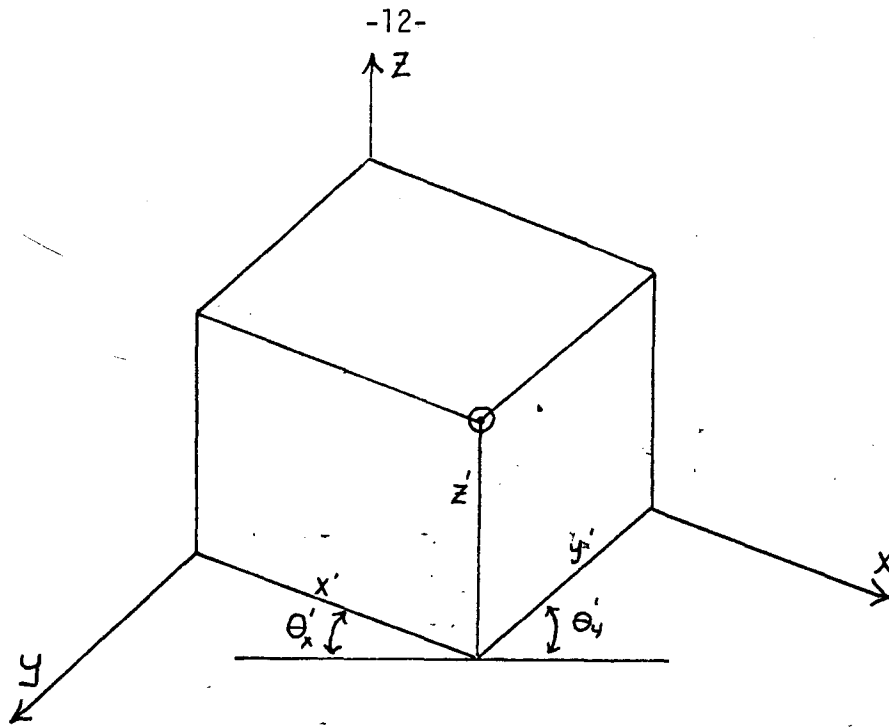
FIG. 1.2.1

TOP VIEW

SIDE VIEW

FIG. 1.2.2

We also know that Θ_x plus Θ_y must equal $\pi/2$ which gives us:

$$\Theta_x + \Theta_y = \pi/2$$

$$\tan \Theta_x = \tan (\Theta_x + \Theta\pi/2)$$

$$\tan \Theta_x = \frac{1}{\tan \Theta_y} \qquad (3)$$

We now have three equations which we can solve for our three unknowns: Θ_x, Θ_y and S. Solving them we get:

$$S = \sin^{-1} \left(\sqrt{\tan \Theta_x' \ \tan \Theta_y'} \right) \qquad (4)$$

$$\Theta_x = \tan^{-1} \left(\sqrt{\frac{\tan \Theta_x'}{\tan \Theta_y'}} \right) \qquad (5)$$

$$\Theta_y = \tan^{-1} \left(\sqrt{\frac{\tan \Theta_y'}{\tan \Theta_x'}} \right) \qquad (6)$$

THE FORESHORTENING OF THE THREE DIMENSIONS

Determining the foreshorrtening of Z is easy. We already know that the angle between Z and our image plane is S so that its apparent length will be equal to its true length times the cosine of S (Refer

to figure 1.2.3).

$$Z = \frac{Z}{\cos S} \qquad\qquad (7)$$

Determining the true length of the other two dimensions is a little

more difficult.   Figure 1.2.4 shows the top surface of our block.

We resolve X into components α and β where α is parallel to the horizontal

line and β is perpendicular to the horizontal line.

From figure 1.2.4A we have:

$$\alpha' = X' \cos \Theta'_X$$

$$\beta' = X' \sin \Theta'_X$$

From figure 1.2.4B we have:

$$X^2 = \alpha^2 + \beta^2$$

Since α is parallel to our image plane, its projection α' will

appear in it's true length.   We know that β intersects the image plane

at an angle $\pi/2$ - S so that:

$$\beta = \frac{\beta'}{\sin S}$$

(Refer to figure 1.2.5)

So we get:

$$X^2 = \alpha^2 + \beta^2$$

$$= \alpha'^2 + \frac{\beta'}{\sin S}^2$$

$$= X'^2 \cos^2 \Theta'_X + \frac{X'^2 \sin^2 \Theta'_X}{\sin^2 S}$$

$$X = X' \left[\cos^2 \Theta'_X + \frac{\sin^2 \Theta'_X}{\sin^2 S} \right]^{1/2} \qquad\qquad (8)$$

Similarly:

$$Y = Y' \left[\cos^2 \Theta_y' + \frac{\sin^2 \Theta_y'}{\sin^2 S} \right]^{1/2} \qquad (9)$$
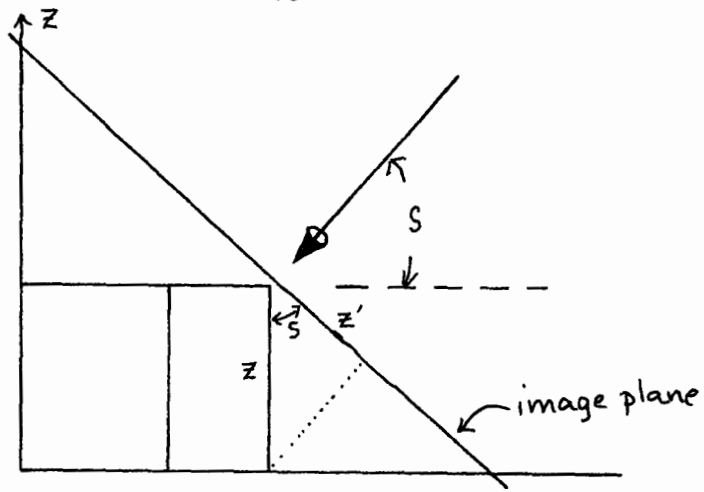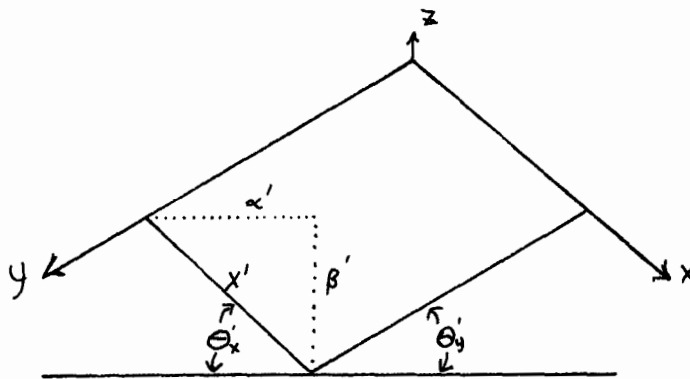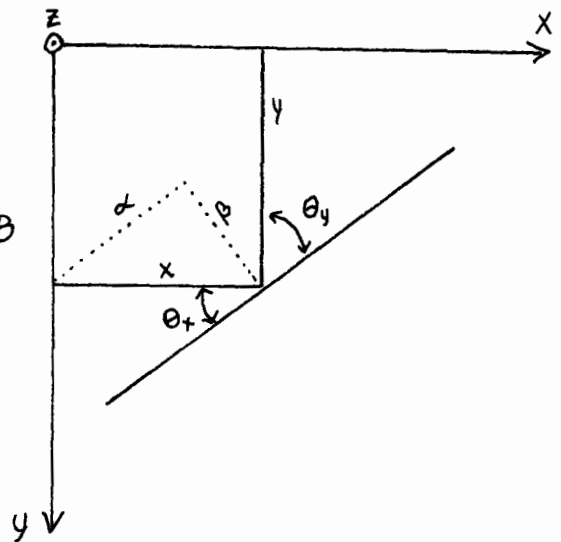
FIG. 1.2.3

OBLIQUE VIEW

FIG. 1.2.4 A

FIG. 1.2.4

Top View

FIG. 1.2.4 B

Let's see how this works for an example, the scene in figure 1.2.6.   We have:

$$\Theta_x' = 17°$$

$$\Theta_y' = 25°$$

$$X' = 5.2 \text{ cm.}$$

$$Y' = 4.5 \text{ cm.}$$

$$Z' = 6 \text{ cm.}$$

And we calculate:

$$S = \sin^{-1}\left(\sqrt{\tan 17° \, \tan 25°}\,\right)$$

$$= 22°$$

$$X = 5.2 \text{ cm.} \left[\cos^2 17° + \frac{\sin^2 17°}{\sin^2 22°}\right]^{1/2}$$

$$= 6.4 \text{ cm.}$$

$$Y = 4.5 \text{ cm.} \left[\cos^2 25° + \frac{\sin^2 25°}{\sin^2 22°}\right]^{1/2}$$

$$= 6.5 \text{ cm.}$$

$$Z = \frac{6 \text{ cm.}}{\cos 22°} = 6.5 \text{ cm.}$$

Thus the object is actually a cube.

FIG. 1.2.5

FIG. 1.2.6

1.3 Identifying the Shape of Some Objects Through Their Dimensions

Once we have determined the true dimensions of an object, we can use this information to help us identify its shape. For example, we could divide the class of rectangular parallelepiped into cubes, bricks, boards, sticks, sheets, etc. as in figure 1.3.1.

Which of the shapes in figure 1.3.1 an object might best match would depend on its relative dimensions. One means of identifying such an object would be to look at the ratio of its longest dimension to its next-to-longest dimension and the ratio of its next-to-longest dimension to its shortest dimension. With these ratios we could define a sort of two-dimensional object space as in the graph in figure 1.3.2.
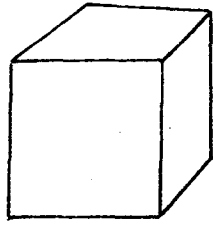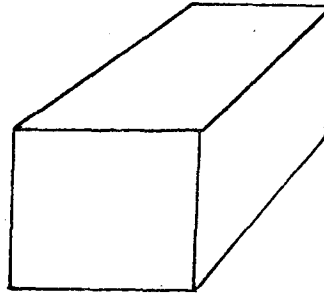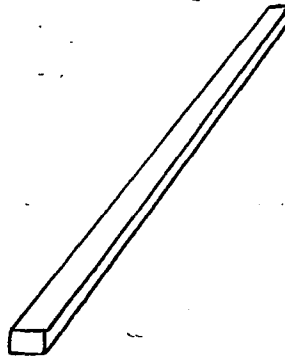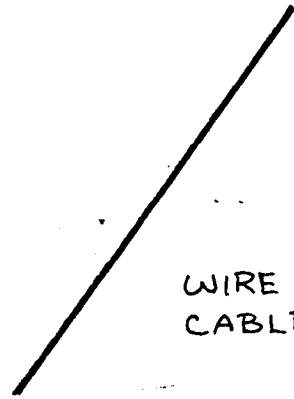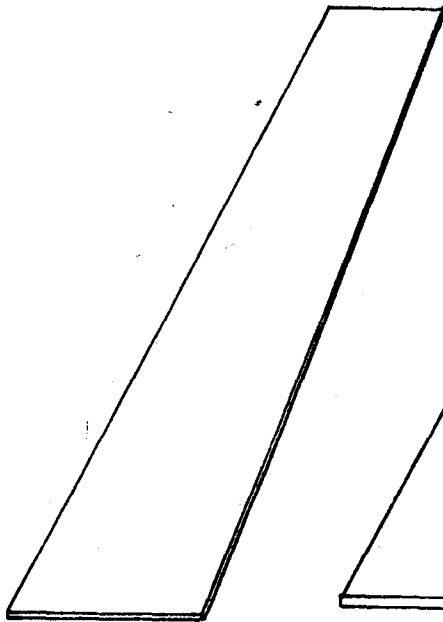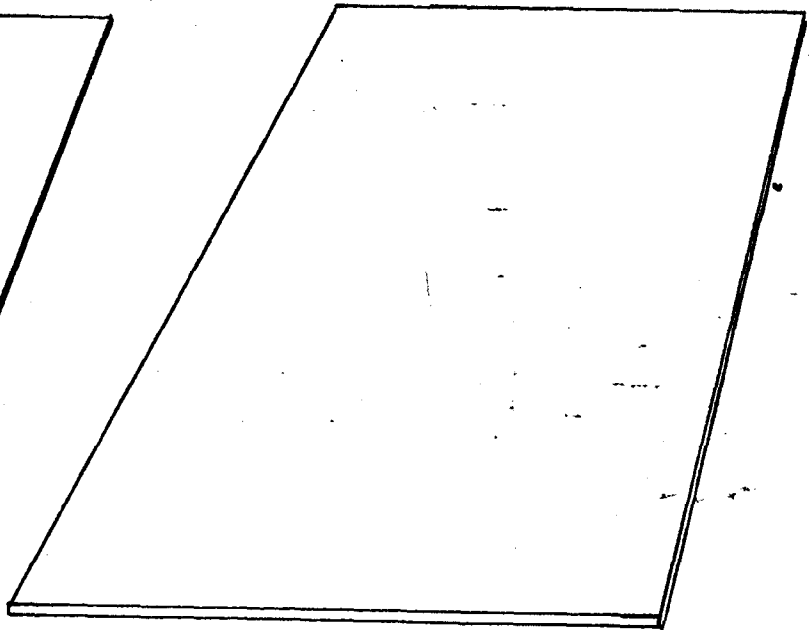
CUBE   BRICK
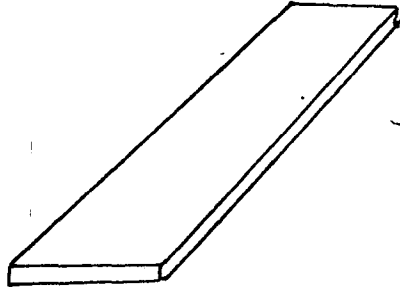
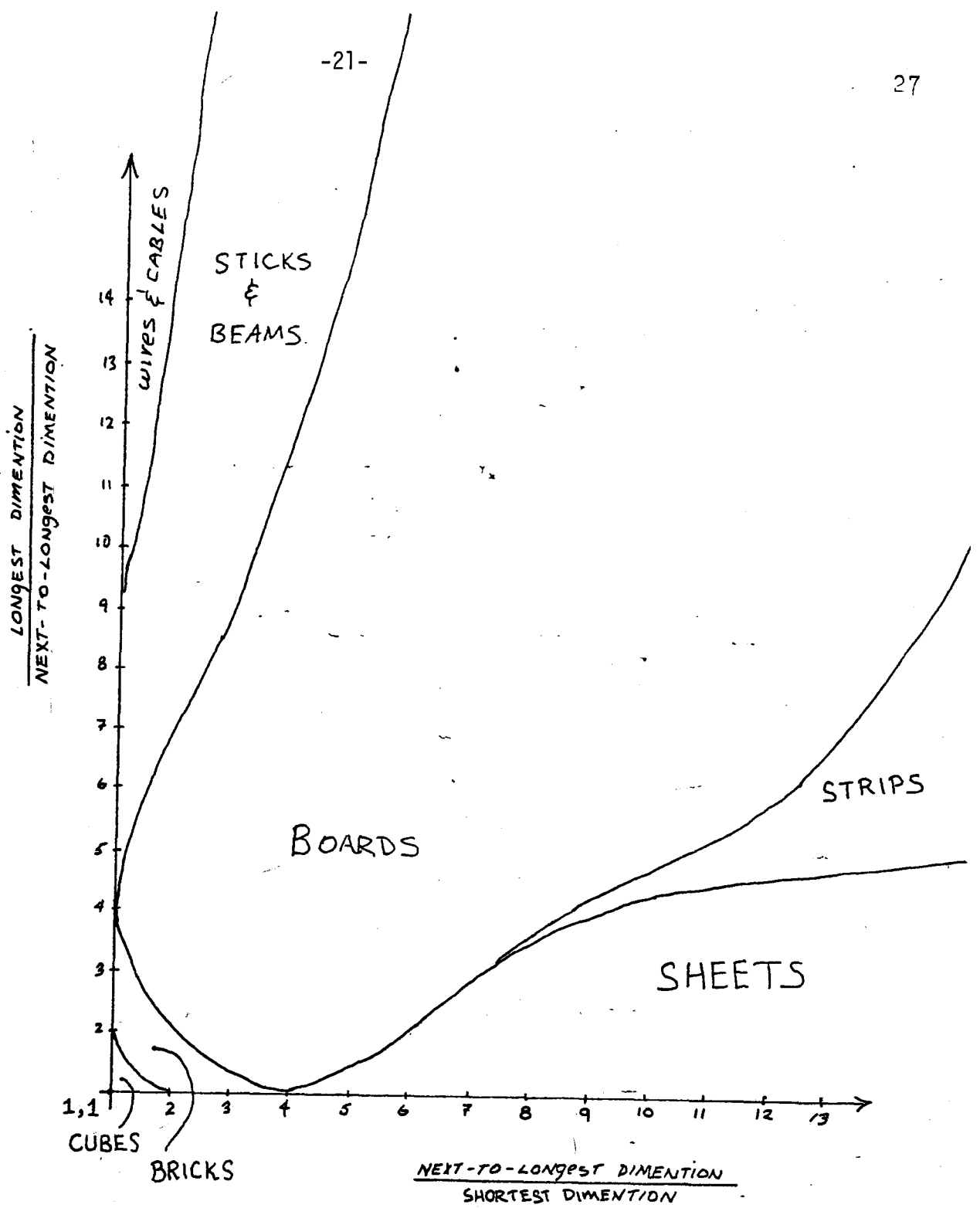BOARD   STICK or BEAM   WIRE OR CABLE

STRIP   SHEET

FIG. 1.3.1

Rectangular Parallelepiped Space

FIG. 1.3.2

3. THE IN-FRONT-OF RELATION

3.0

The IN-FRONT-OF relation is a useful one in developing descriptions of scenes. Winston has developed an algorithm for detecting the IN-FRONT-OF relation which works well except for cases where the objects are aligned. The method works by noting that if one of two objects appears to obscure the other, then it is either above or in front of the other.

Obscuring objects are detected by searching for particular types of T-joints on the periphery of the objects. The type of T-joints sought are those whose crossbars belongs to the object conjectured to be the obscuror. If both regions bordering the shaft of the T belong to a second object, or if one shaft-bordering region belongsto a second object and the other to the background, then the object belonging to the crossbar of the T-joint obscures that second object. Figure 3.0.1 shows both kinds of qualifying T-joints. This yields pairs of obscuring and obscured objects which are candidates for the IN-FRONT-OF relation. A separate program which determines when one object is above another (and works by a different method) is then called to reject all candidates for which the ABOVE relation is known to hold.

The weakness of this algorithm is that the seam between the obscuring and obscured objects may not exhibit the required type of T-joints. This is the case when the objects are aligned, as in figure 3.0.2. This is a serious problem since aligned objects are frequently encountered, no doubt
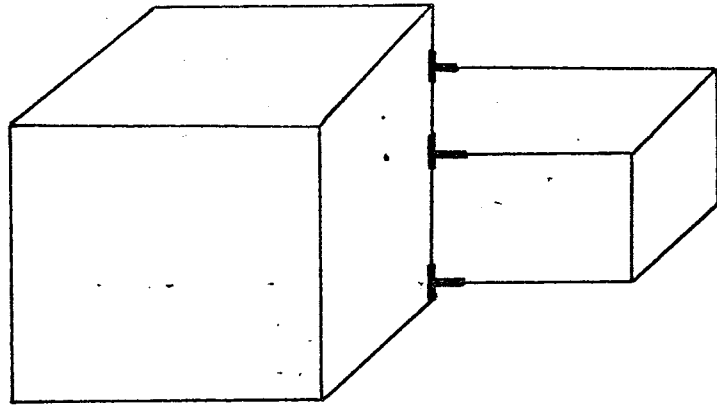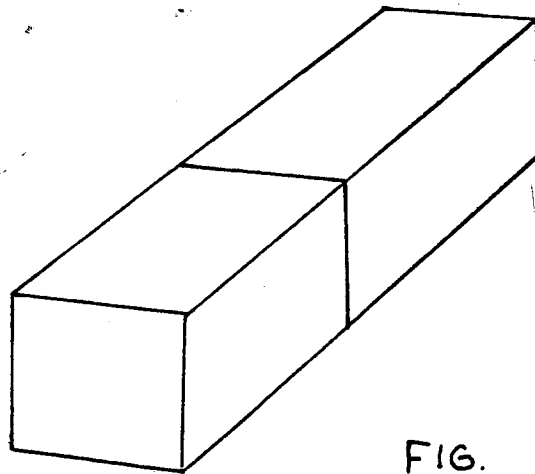
FIG. 3.0.1

FIG. 3.0.2

due to some human predilection to order.

3.1 The IN-FRONT-OF Relation for Aligned Objects

Objects which meet at X vertices, K vertices or certain types of T vertices suggest alignment and resulting occlusion of one object by another.   Figure 3.1.1 shows these cases.   To determine the IN-FRONT-OF relation for aligned objects we can use a two part procedure similar to Winston's.   First, examine the X, K, and T vertices on the seam between two objects to find pairs of obscuring and obscured objects which are candidates for the IN-FRONT-OF relation.   Second, we use the program which detects the ABOVE relation to reject all candidates for which that relation is known to hold.

If two objects meet at an X vertex (where only two of the lines are collinear) then the object on the narrow angle side of the non-collinear lines is the obscuring object.   See figure 3.1.2.

If two objects meet at a K vertex, then the object which has one of the legs of the K as its interior line is the obscuring object, as in figure 3.1.3.
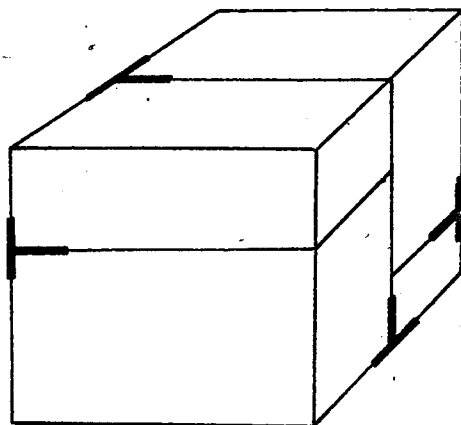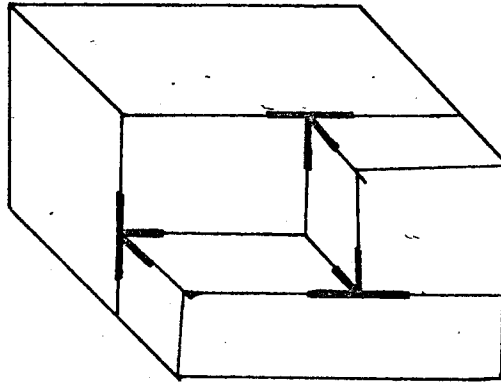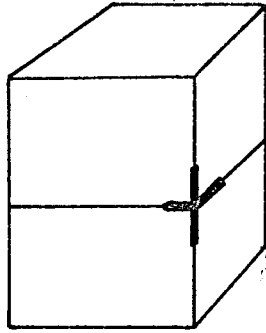
ALIGNED OBJECTS

FIG. 3.1.1

If two objects are separated by the stem of a T vertex with the background on the wide angle side of the T, then the object on the side of the larger of the angles formed by the stem and the crossbar of the T, obscures the other object. In figure 3.1.4 we can tell that object A obscures B by virtue of vertex 1, A obscures C by vertex 2, B obscures D by vertex 3, and C obscures D by vertex 4. This type of T vertex is actually a degenerate form of X vertex, since these T's would be seen as X's if viewed from other angles.

A K vertex, when viewed from some angles, can also appear as a T vertex as is shown in figure 3.1.5. A T vertex of this type will have its wide angle side and one other side belonging to one object and the other side belonging to a second object. If we have such a T vertex formed by two objects, then the object belonging to only one side of the T vertex will be the obscuring object. In figure 3.1.6 this heuristic is necessary in determining that object A obscures object C.

These four heuristics will suggest pairs of objects which are candidates for the IN-FRONT-OF relation. Part two of the procedure rejects all candidates for which the ABOVE relation is known to hold.
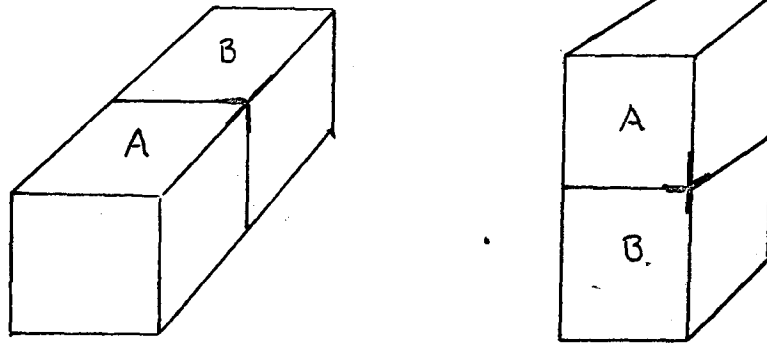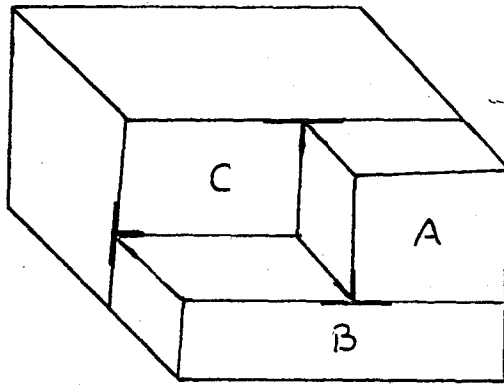
A OBSCURES B

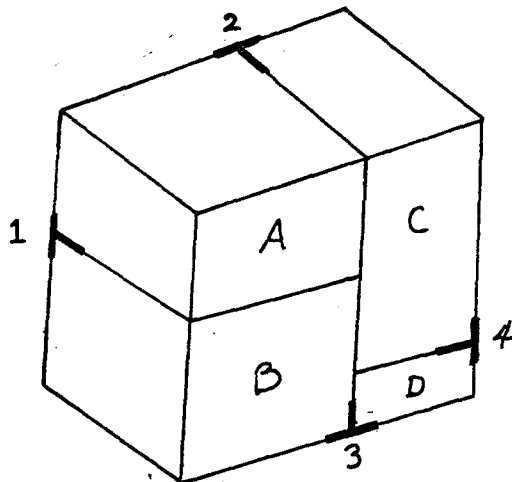FIG 3.1.2

FIG. 3.1.3

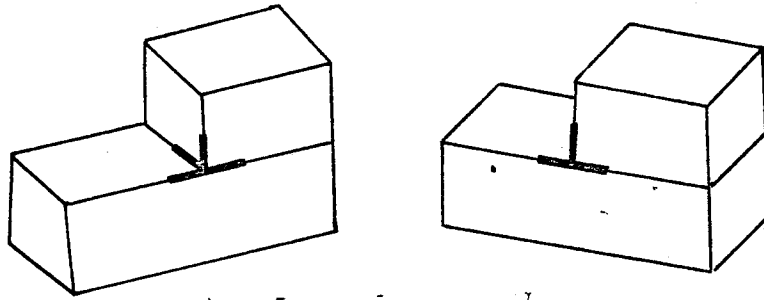A OBSCURES B & C

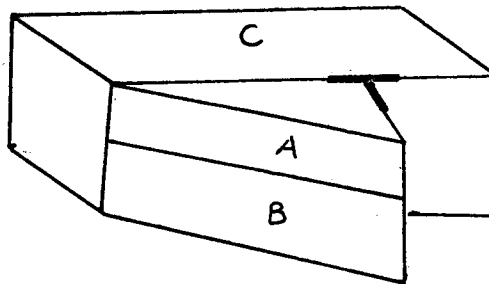B OBSCURES C

FIG. 3.1.4

FIG. 3.1.5

FIG. 3.1.6

3.2 An Assumption And A Problem

This algorithm uses Winston's procedure for determining ABOVE
relationships to reject improper candidates for the IN-FRONT-OF relation.
Winston's procedure assumes that we view the scene from above. If we knew
that we were viewing the scene from below, then we could use a procedure
which determined if one object were below another to filter our candidates
for the IN-FRONT-OF relation. In general, we could check both. That
is, if object A is known to obscure object B and A is neither above nor
below B, then A is in front of B.

The heuristics which examine X-vertices and certain T-vertices to
discover obscured and obscuring objects work mainly for objects whose edges
meet at right angles where the X- and T-vertices are created. If this is
not the case, then these heuristics will make mistakes. For example,
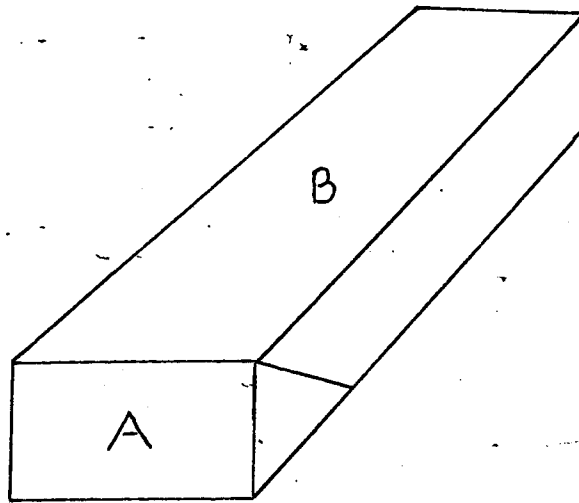in figure 3.2.1, the T-vertex suggests that B is in-front-of A, since B is
not above A.

FIG. 3.2.1

3.3 What Does IN-FRONT-OF Mean Anyway?

The definition of IN-FRONT-OF that we have been using is that if one object obscures a second object but is not above it, then the first object is in-front-of the second.   This definition does not seem to be in complete harmony with the human notion of IN-FRONT-OF.   Consider the scene in figure 3.3.1.   By our definition, object B is in-front-of object A, whereas humans would probably prefer the reverse or noncommitment.

Again, consider the sequence of drawings in figure 3.3.2.   In this sequence, when does object B become in-front-of object A?   By our definition in scene two, as soon as object B obscures object A.   In my own judgment, B becomes in-front-of A somewhere around scene five.   The natural orientation of the objects seems to have a strong influence.   To me, scenes one through four exhibit more of a left-right relationship. Even in scene five I am hesitant to say that B is in-front-of A.

Occlusion must certainly play an important part in the concept of IN-FRONT-OF, but only a part.   The fact that humans can manage visual perceptions quite well with only one eye suggests that there is more to be learned without jumping to stereoscopic vision.
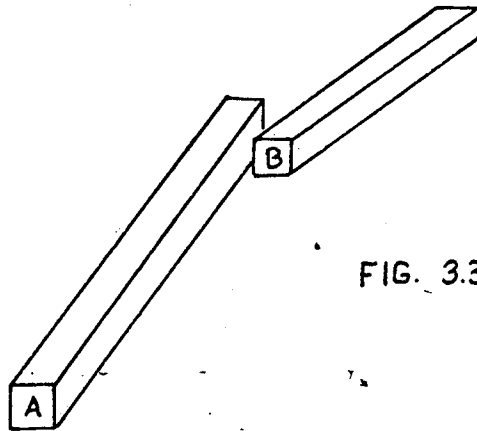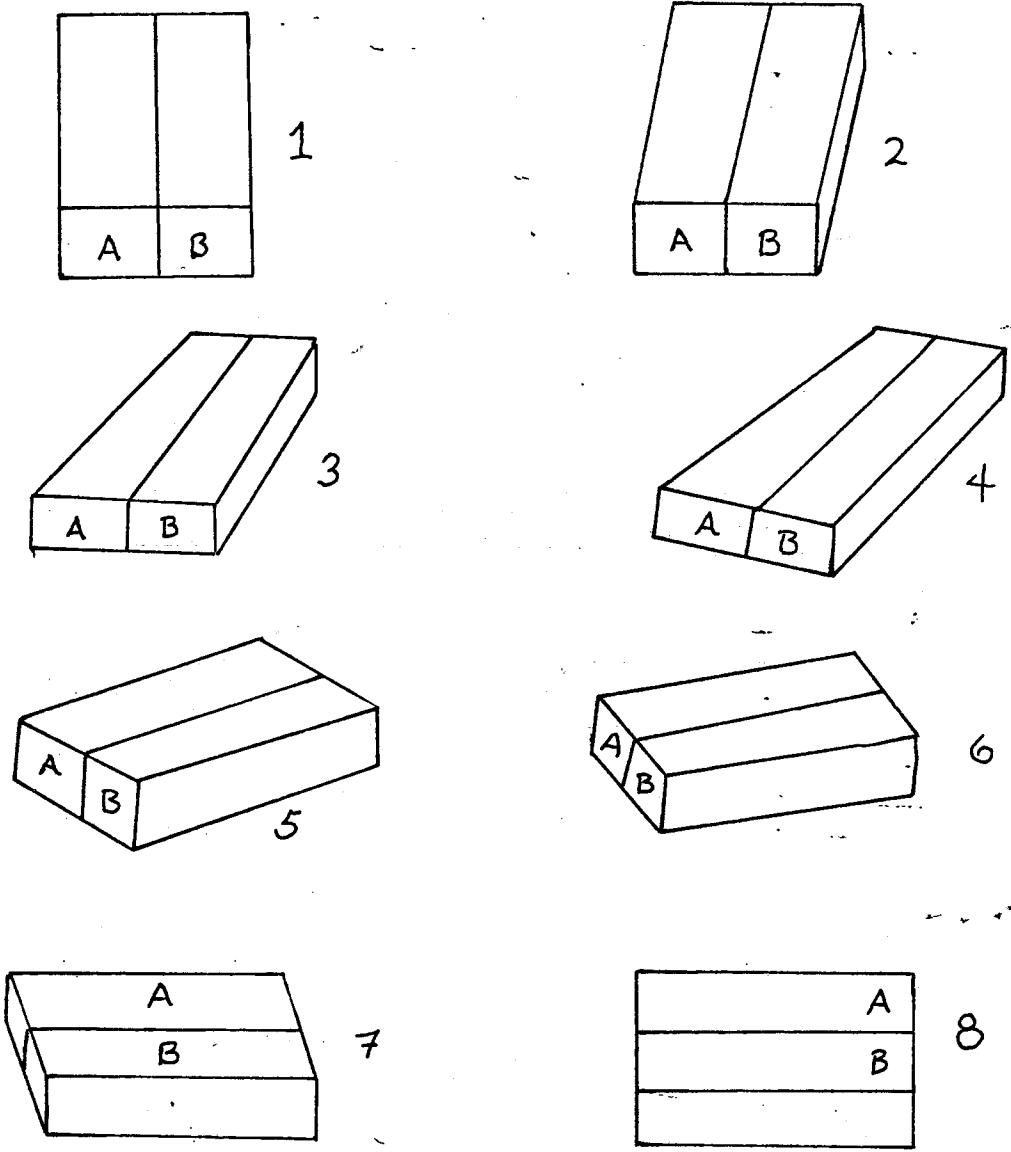
FIG. 3.3.1

FIG. 3.3.2