COGNITIVE SCIENCE 3, 231–250 (1979)

Some Demonstrations of the Effects of Structural Descriptions in Mental Imagery*

GEOFFREY HINTON
University of California, San Diego

A visual imagery task is presented which is beyond the limits of normal human ability, and some of the factors contributing to its difficulty are isolated by comparing the difficulty of related tasks. It is argued that complex objects are assigned hierarchical structural descriptions by being parsed into parts, each of which has its own local system of significant directions. Two quite different schemas for a wire-frame cube are used to illustrate this theory, and some striking perceptual differences to which they give rise are described. The difficulty of certain mental imagery tasks is shown to depend on which of the alternative structural descriptions of an object is used, and this is interpreted as evidence that structural descriptions are an important component of mental images. Finally, it is argued that analog transformations like mental folding involve changing the values of continuous variables in a structural description.

1. INTRODUCTION

Recent experiments by Shepard and his colleagues (Shepard & Metzler, 1971; Cooper & Shepard, 1973; Cooper, 1976) have helped to revive interest in the nature of the mental images we have when we visualize objects. Some authors (Minsky, 1975; Pylyshyn, 1973) have argued that mental images are like the structural descriptions or relational networks that have proved useful for representing syntactic or propositional structures. This has been contested by others (Paivio, 1977; Kosslyn & Pomerantz, 1977) who have stressed the wholistic, analog nature of images.

The stimulating aspect of the experiments by Shepard and his coworkers

*Most of this research was performed in the Cognitive Studies Program, University of Sussex, under SRC Grant No. BRG 8688-7. Requests for reprints should be sent to Geoffrey Hinton, Center for Human Information Processing C-009, University of California, San Diego, La Jolla CA 92093.

was that they used an objective measure (reaction times) to show that when we visualize an object rotating, there is a changing internal representation whose intermediate states correspond to intermediate orientations of the object. However, as Shepard recognizes, this close correspondence between internal and external changes does not demonstrate that the internal representations are spatially isomorphic to external objects. Reaction times are not so helpful with this issue.

This paper attempts to clarify the nature of mental images by analyzing tasks in which people are required to construct or manipulate mental images of geometric structures. Success or failure can be objectively assessed either by asking questions about relationships between parts of the constructed objects, or by making subjects point out, in real 3-D space, a set of consistent locations for specified parts of the structure. This method of externalizing mental images has an advantage over drawing, since it prevents subjects from using perception of what they have already externalized to help them in producing the rest.

The aim of the paper is to present some new demonstrations and to show that they can be interpreted as evidence that mental imagery involves a particular kind of structural description. The approach to the internal representation of spatial structures is far from new (Clowes, 1969; Narisimhan, 1966; Palmer, 1977; Reed, 1974), but its application to the phenomena to be described serves both to refine the basic idea and to corroborate its relevance to human spatial awareness. Baylor (1971) has already shown that a program can solve mental imagery tasks in a psychologically plausible way by manipulating structural descriptions. Baylor studied the Guilford block visualization task which involves mentally slicing a cuboid with planes normal to its principal axes. The mental imagery tasks reported here are different, and the emphasis is on showing the differential effects of alternative structural descriptions for the same object. However, the conclusion is the same: that mental imagery involves the manipulation of structural descriptions.

The tasks described below have been presented informally to many subjects with varied backgrounds. Generally, the difficulties were obvious and the relative difficulties of different tasks were consistent between subjects, so the tasks can be treated as demonstrations, as can the perceptual phenomena described in section 5. The reader is encouraged to try the tasks and to construct a cube from pipe-cleaners in order to observe the perceptual phenomena.

2. A VERY HARD IMAGERY TASK

TASK 1: Imagine a wire-frame cube resting on a tabletop with the front face directly in front of you and perpendicular to your line of sight. Imagine the long diagonal that goes from the bottom, front, left-hand corner to the top, back, right-hand one. Now imagine that the cube is reoriented so that this diagonal is vertical and the cube is resting on one corner. Place one fingertip about a foot above a tabletop and let this mark the position of the top corner on the diagonal. The corner on which the cube is resting is on the tabletop, vertically below your fingertip. With your other hand point to the spatial locations of the other corners of the cube.

2.1 Results

I have tried this experiment (with some minor variations of wording) on over twenty people, many of them research workers in vision. Only one of them produced the correct answer and only a few realized that there were six remaining corners to be accounted for. The rest simply assumed that there were four remaining corners and frequently placed them in a square lying in a horizontal plane halfway between the top and bottom corners. This may suggest that the subjects thought of the cube as resting on one edge, but they claimed that they did not. Many subjects took very much longer than a few seconds to produce an answer and when they did they were clearly not confident about it.

Figure 1 shows that the six remaining corners form a zigzag ring in which

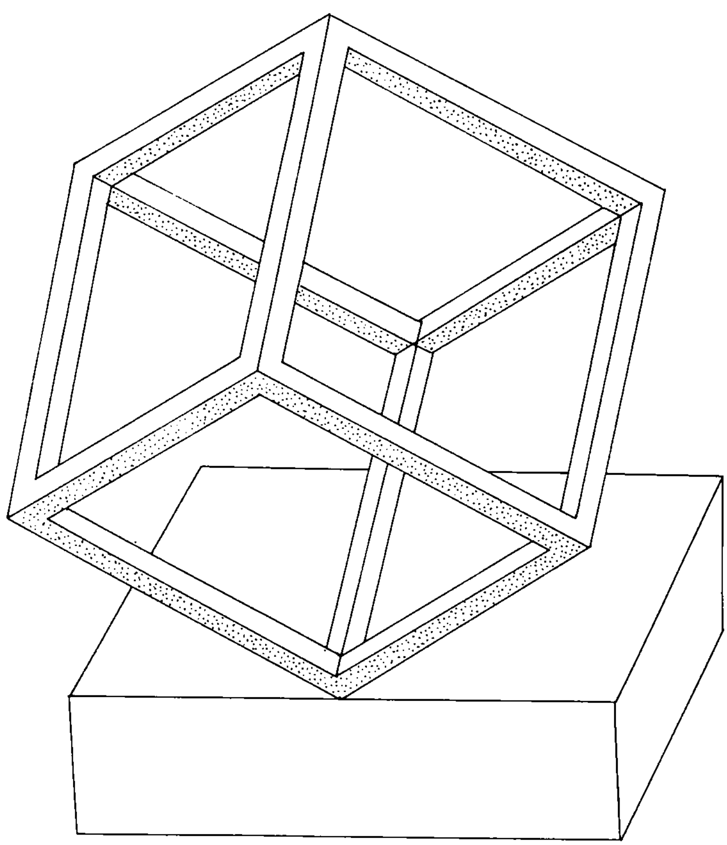


Figure 1 A wire-frame cube resting on one corner with the diametrically opposed corner vertically above it. The edges are depicted as rectangular beams to increase the impression of depth. The rectangularity of the cube itself, which normally helps to convey depth, is not effective when the cube is interpreted as having the vertical axis shown here.

the corners are alternately one third and two thirds of the height of the vertical axis.

3. WHY THE TASK IS HARD

Several different aspects of the task contribute to its difficulty:

1. The diagonal which must be made vertical has components in all three of the mutually orthogonal directions defined by the initial position of the cube.
2. Perceptual examination of a stationary cube in the initial, standard orientation is not allowed during the task.
3. The correct response requires a detailed awareness of the internal structure of the cube and this structure is fairly complex involving eight corners linked by twelve edges.

In the rest of this section, each of these factors is isolated and shown to be a cause of difficulty by comparing similar tasks which differ on only one factor.

3.1. The Effect of the Direction That Must Be Made Vertical

Task 1 can be compared with one in which the direction which defines the new vertical has components in only two of the three orthogonal directions defined by the cube. An obvious comparison task involves rotating the cube so that a diagonal of the front face becomes vertical. Now two of the corners remain on the tabletop and the task seems easy, perhaps because most subjects know that a square which has been tilted so that its diagonal is vertical is equivalent to an upright diamond. The availability of the diamond schema is a confusing factor, so a different task was chosen instead.

> TASK 2: Imagine a cube sitting squarely in front of you on a tabletop, as before, and also imagine a line which lies in the front face and joins its bottom left-hand corner to the middle of its right-hand, vertical edge. Picture this line clearly in your mind. Now imagine rotating the cube counterclockwise so that the line becomes vertical. Place a fingertip about a foot above a tabletop to indicate the top of the line, and with the other hand, point to the corners of the cube.

Most of my subjects were unable to imagine the whole cube rotating, but they were able to work out where the corners would end up. They were thus better at task 2 than at task 1. They all noticed that the back face would remain aligned with the front one, and therefore concentrated on the subproblem of finding the new positions of the front corners. It appears that task 1 is more difficult than task 2 because it cannot be reduced to a simpler 2-D subproblem in the same way.

3.2. The Effect of Perceiving the Object

If a wire-frame cube sitting squarely on a tabletop is examined with task 1 in mind, certain facts are immediately obvious. For example, the corners which do not lie on the specified long diagonal are not coplanar, and there are more than four of them. These facts were far from obvious to most of the subjects who were only allowed to imagine the cube. So it seems that the mental image of a cube does not explicitly represent some of the relationships that are perceptually available, nor does it allow them to be easily inferred. This observation may appear to conflict with the claim by Podgorny and Shepard (1978) that mental images are very similar to percepts. However, their view is quite compatible with the evidence if one allows that there may be several different ways of representing the spatial structure of an object, and that at any moment only one representation can be entertained. Perception allows us to switch from one percept to a quite different one for the same external structure, whereas imagery only seems to allow this in familiar cases (square/diamond) or when the structure is rather simple, like an N which can be mentally rotated through 90 degrees and reimagined as an upright Z. Later, I describe further cases in which it is hard to alter the way in which an imagined spatial structure is represented.

3.3. The Effect of Complexity

TASK 3: Look at the structure formed by the three heavy edges in figure 2. Now, without looking at the figure again, imagine this three-limbed structure rotating so that the top, back, right-hand free end is vertically above the other free end. Place a fingertip about a foot above a tabletop to mark the position of the top free end, and point with the other hand to the positions of the corners between the limbs.

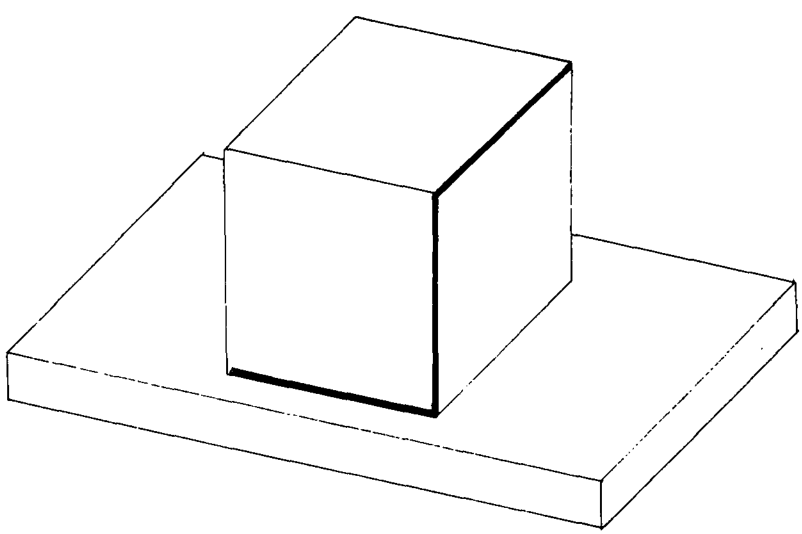


Figure 2 A structure of three mutally perpendicular limbs, shown embedded in a cube.

Subjects who had not attempted task 1 found this task difficult even though they were shown a real three-limbed object which was removed before they were told the task. Some gave incorrect responses, and most of those who succeeded reported that they mentally reconstructed the object in the final orientation rather than imagining the rotation. Nevertheless, subjects were markedly better at task 3 than at task 1. Since the tasks involve just the same rotation and differ only in the complexity of the structure (as measured by the numbers of edges and corners) it could be argued that the complexity of the cube is one source of difficulty in task 1. However, complexity is a problematic concept in this context. The idea that it can be measured by the numbers of corners and edges conflicts with the observation that many people find the whole cube in figure 2 easier to imagine than just the three heavy edges.

A representation which uses some of the symmetries of the cube to avoid duplicate data structures for symmetrical parts (Leeuwenberg, 1971; Shneier, 1978), could explain why the whole cube appears so simple in its normal orientation, and why its apparent complexity increases when these symmetries are no longer helpful, as in task 1. The psychologically relevant measure of complexity would then depend on how the cube was represented rather than on its objective properties alone. This would resolve the apparent paradox that the relative complexities of the cube and the three-limbed object in figure 2 appear to change when task 1 is considered. Inappropriate concentration on properties of the cube, like its square cross-sections, which normally make it simple, might well explain the common wrong answer that there are four remaining corners forming a horizontal square.

Any comparison of task 1 with task 3 must also consider a further factor. Even if the new positions of the individual corners were as easy to locate in task 1 as in task 3, task 1 would still have the extra response requirement of relating the six corners to one another. We shall return to this issue in section 7.

4. ASSIGNED SIGNIFICANT DIRECTIONS AND SPATIAL REPRESENTATIONS

The difficulty of task 1 emphasizes the difference between two ways of organizing the elements of a cube. Normally a wire-frame cube is seen as having edges which are parallel or perpendicular to an intrinsic top/bottom direction. If the cube is tilted, its intrinsic top/bottom direction tilts with it, so that the orientations of the edges relative to this direction remain the same. This allows the spatial structure to be given a constant internal representation whatever its orientation, provided the top/bottom direction can be assigned appropriately. Rock (1973) has provided a considerable amount of evidence showing that the intrinsic top/bottom directions assigned to 2-D objects lying in the frontal plane affects their phenomenal shape and hence their recognition.

A consequence of achieving orientation invariant descriptions by describing structures relative to assigned intrinsic directions is that a single structure may have several quite different descriptions if there are several different, sensible ways of assigning intrinsic directions to it. The tilted-square/upright-diamond ambiguity is a familiar example. I shall call this a type 2 ambiguity to distinguish it from the type 1 ambiguities exhibited by the Necker cube or the Rubin vase/faces picture (Gregory, 1966). In type 1 ambiguities, there are alternative internal representations of a single picture or retinal image which correspond to different external structures (e.g., the orientation of the Necker cube appears to change when it flips), whereas in type 2 ambiguities, the alternative internal representations correspond to the same external structure. A linguistic example of a type 2 ambiguity is "Next weekend we shall be visiting relatives." "Visiting" can be interpreted as an adjective or a verb and this leads to two quite different senses (internal representations) with the same truth conditions. Type 2 ambiguities are particularly good for revealing the nature of our representations. They allow us to see how our awareness of something is affected by the way we represent it, since for both representations the external entity is the same.

A strikingly different internal representation of the spatial structure of a cube is obtained when a diagonal through its center is used to define the intrinsic top/bottom direction. This way of perceiving a cube is easier when the diagonal aligns with the contextual vertical provided by gravity, the room, or the picture frame as in figure 1. The edges of the cube can then be seen to form two tripods, an upright one at the top and an inverted one at the bottom. The tripods are rotated 60 degrees relative to each other about the vertical axis and their feet are joined by a "horizontal" zigzag ring. The corners in this ring are alternately one third and two thirds of the height of the vertical axis. For convenience, this way of representing a cube will be referred to as the "hexahedron" schema, and the more familiar way will be called the "cube" schema.

4.1. Significant Directions and Hierarchical Organization

A number of authors (Minsky & Papert, 1972; Turner, 1974; Palmer, 1975; Marr & Nishihara, 1978) have argued that complex spatial structures are represented as a hierarchy of parts, each of which is defined in terms of lower level parts and their spatial relationships. Marr and Nishihara distinguish two ways of assigning significant directions to such hierarchical structures. They point out that the orientations of parts and the spatial relations between them need not all be represented relative to a global system of significant directions. Instead, each part may have its own intrinsic top/bottom direction which is used for representing the spatial relations between the immediate constituents of that part. For example, a person's eyes are seen to lie on a "horizontal" relative to the top/bottom direction defined by his face, even when he has his head on one side.

4.2. Local Coordinate Systems

Marr and Nishihara actually claim that for many objects, particularly biological ones, each part has its own associated cylindrical coordinate system, relative to which the constituents of the part are located. This is somewhat stronger than the claim that significant directions such as top/bottom and front/back are locally defined, since it also involves choosing a location for the origin of the local coordinate system. There are reasons for doubting that people commit themselves to a full coordinate system with its local origin. The evidence that people represent spatial structures relative to assigned directions is that different assignments lead to phenomenally different representations of the same structure. If we also chose origins and used full coordinate systems, then different choices of origin should have produced phenomenally different representations even when the directions were the same. I have been unable to find any evidence of this phenomenon.

It is quite feasible to represent spatial structures by representing spatial relations between pairs of parts without relating each part to a single special location like the origin (see Kuipers, 1978 for an example). The need to choose origins and significant directions depends on the number of parts that are involved in each of the relationships that are used to characterize a spatial structure. If the locations of the individual parts are used separately, these locations must be described relative to an origin and assigned directions. If pairs of parts are used it is unnecessary to relate each part to an origin, but it is still necessary to have assigned directions. If triples of parts are used, the angles formed by three parts can be represented without any need for the assigned directions on which people seem to depend.

One argument in favor of the use of full coordinate systems is that each part must be given a location in order to describe its relations to other parts, so this location could act as the origin when a local coordinate system is needed for describing the internal structure of a part. This argument is suspect because the relationships between parts can be roughly described using locations for parts which are not nearly accurate enough to act as local origins for obtaining cannonical descriptions of the internal structures of the parts.

4.3. Orthogonal Direction Systems

For those familiar with Cartesian coordinate systems, it seems sensible to use a system of three mutually perpendicular directions, and natural language supports this inclination by providing the terms top/bottom, front/back, and left/right, all of which can be used to refer to as object's intrinsic directions (Miller & Johnson-Laird, 1976). Several phenomena show that orthogonal systems of directions are important for our perception of spatial structures as well as for our conscious mathematical and verbal descriptions of them. It is clear that we have a strong tendency to interpret line drawings (e.g., the Necker cube) or real images (e.g. the image produced by the Ames room) as depicting objects with rectangular corners, provided this is geometrically possible (see Perkins, 1976), even though such drawings or images could perfectly well depict nonrectangular objects like the actual Ames room.

There are two rather different views about the role of orthogonal systems of directions in perception:

1. In our carpentered world, orthogonal direction systems are often helpful and we use them in our spatial representations when appropriate.
2. Orthogonal direction systems are fundamental to our higher level spatial representations. We always represent orientations relative to such systems, even for thoroughly nonrectangular objects.

The difference between the first and second views can be seen when the hexahedron schema is considered. The choice of top/bottom direction in this schema prevents any edges from aligning with an orthogonal direction system. So, given the first view, such a system should be abandoned, possibly in favor of a representation based on the threefold rotational symmetry of the object. The second view, however, suggests that the orientations of the edges can only be fully represented if a front/back direction is chosen, and so on this view a mental image of a hexahedron is incomplete if no front/back direction is assigned. The way we represent threefold rotational symmetry (e.g., an equilateral triangle) is a crucial test. Evidence that it is represented in terms of one or other of the three bilateral symmetries (Attneave, 1968) favors the second view.

4.4. Viewer-Centered Representations

The direction systems mentioned so far have been based on objects or their parts. They are useful because they yield representations of spatial structures which are independent of the viewpoint. Such representations provide a stable world which does not change as we move around, and they are also useful for recognizing objects. However, their relationship to the retinal image is complex, and a viewer-centered system of directions is much more convenient for early processing. This system could be polar, using a distance from the retina plus two angles corresponding to the retinal vertical and horizontal. Alternatively, locally parallel systems could be used (Ullman, 1977) so that for any local region of space there were orthogonal directions defined by the retinal vertical and horizontal and the line of sight. Sources of depth information like occlusion, stereo, and shape from shading may well be integrated in a viewer-centered representation (Marr, 1978).

If visual imagery involves having representations of the kind normally generated by perception, viewer-centered representations may form a part of mental images. However, it will be argued that mental images must also involve representations that use object-based direction systems, and it is this latter component of images that will be the focus of interest.

5. ALTERNATIVE SCHEMAS FOR A CUBE AND THEIR PERCEPTUAL EFFECTS

Comparison of the cube and hexahedron schemas for a wire-frame object shows how different choices of significant directions may correlate with quite different parsings of a structure into parts. The hexahedron (see figure 1) is seen as having three edges at the top, three at the bottom, and six in the middle. A wire-frame cube is often seen as having four edges at the top and four at the bottom, with four parallel edges joining these two groups, though for some purposes (e.g., task 2 in section 3.1) this schema of two squares linked by four parallel edges can be applied differently so that the edges of the front and back faces are picked out as two separate groups. Figure 3a illustrates the higher levels of the cube and hexahedron schemas, and figure 3b shows a possible representation of one part of the hexahedron schema in more detail.

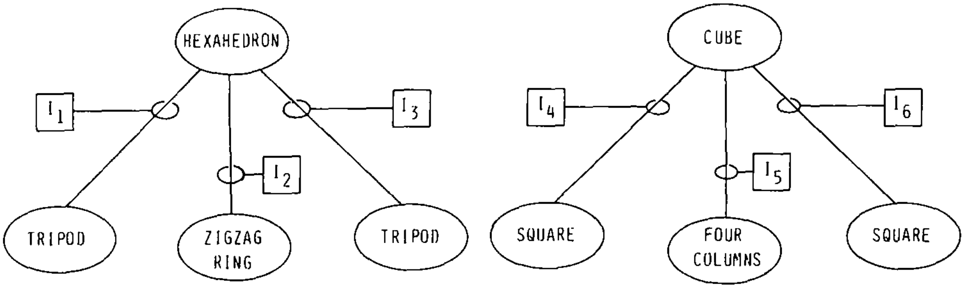
It is not clear that all the edges of a cube need to be represented in working memory when a person imagines a whole cube. It may be that a subject who is familiar with a schema only generates the details when they are needed. Introspection cannot distinguish between details which are always present and details that are rapidly generated as soon as they are looked for.

5.1. Interactions Between Color and Parsing

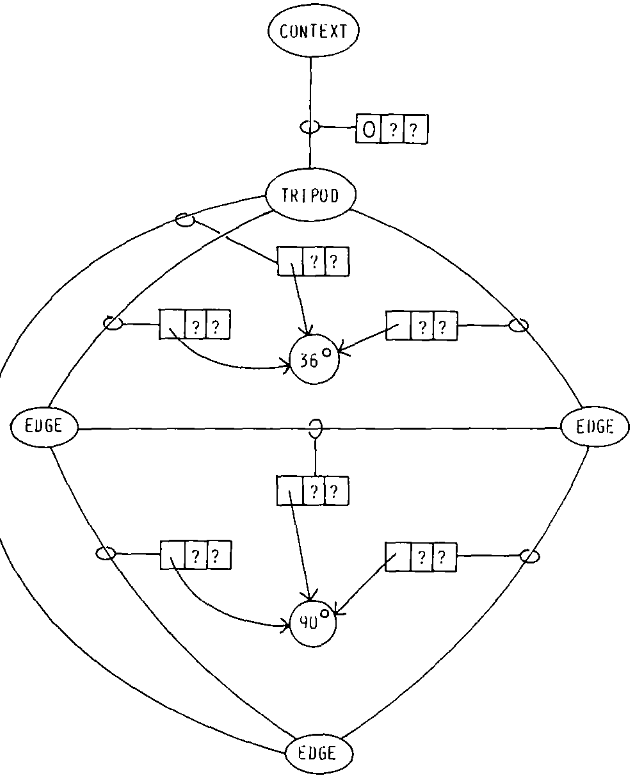
Ideas about the way we parse complex structures can be tested by coloring their constituents. For example, one color can be used for the six edges which form the central ring of a hexahedron, and another color for the rest. The coloring biases people who are familiar with the hexahedron schema toward perceiving the object as a hexahedron, though the bias is weaker than that produced by appropriate orientation. When seen as a hexahedron the coloring appears simple and is easily recalled. By contrast, when the same object in the same orientation is seen as a cube, the colors appear jumbled and disorganized, and if the object is removed, subjects are poor at recalling which edge was which color. The obvious conclusion is that a coloring appears sensible and is easily recalled when it corresponds to the way we parse the object into parts, so that each part is uniformly colored.

5.2. The Effects of Distortion

One technique for discovering what information is explicitly represented within a schema is to observe the effects of various kinds of distortion of the 3-D structure. If a pin-jointed wire-frame cube is perceived as a hexahedron, then elongating the central axis does not seem to affect the "goodness of form" of the object (compare figures 4c and 4d). The perfect cube does not stand out as a specially good case. If, however, the same transformations are perceived using the cube



(a)



(b)

Figure 3(a) The higher levels of two alternative structural descriptions for a cube. Each arc is labelled with the relationship between the intrinsic direction systems of a whole and a part. (b) A representation which uses the threefold rotational symmetry of a tripod. Many of the relationships between the relevant intrinsic direction systems can be specified using just two numerical variables. Each intrinsic relationship is specified by the angles between corresponding significant directions in the two direction systems being related. In this example, only the top/bottom directions have been specified for the tripod and for each edge, so the angles between corresponding pairs of the other significant directions are unspecified.

schema, there is clearly a special case when the angles are right angles, and the other cases are seen as deformed versions of this one (compare figures 4a and 4b). Conversely, if the lengths of the four vertical edges of a cube are varied slightly without changing any angles the cube does not stand out as much more regular than the other cuboids. If, however, this transformation is observed using the hexahedron schema, there is such a pronounced loss of regularity that it is hard to continue using the same schema.

This demonstration shows that different relations are crucial in different schemas. For example, quite different sets of symmetries are made apparent by the cube and hexahedron schemas, so that which transformations preserve perceived symmetry will depend on which schema is used.

The effects of distortion are relevant to an important issue about rectangularity. The Ames room and Necker cube examples appear to demonstrate that interpretations which contain edges joining at right angles are in some sense better than alternative interpretations. If this were so, one would expect that a deformation which destroyed right angles would make a structure look less regular. Indeed, when one looks at a cube being deformed this is just what happens. However, loss of rectangularity does not seem to be perceptually significant when the heaxahedron schema is used. This suggests that it may not be right angles, per se, which are important, but rather the alignment of edges or faces with assigned, orthogonal, intrinsic directions. It is this property of alignment that is lost when the cube is distorted but cannot be lost from the hexahedron because it is never present.

6. NOTICING PARALLELISM IN MENTAL IMAGES

Different descriptions of a spatial structure make explicit different facts about it. The different symmetries made apparent by the cube and hexahedron schemas have already been mentioned. Another striking example is demonstrated by the following task: imagine a hexahedron with its axis vertical. Which edges now have a steeper slope (i.e., are more nearly vertical), the edges in the top tripod or the edges in the central zigzag ring? Subjects who were familiar with the hexahedron schema found this question difficult and often thought the top edges were steeper. In fact, each edge in the top tripod is parallel to two in the zigzag ring, so both sets of edges have the same slope. It appears that facts about which edges are parallel are not readily available from the mental image when the hexahedron schema is being used. Further evidence for this claim is that many subjects are not sure whether an edge in the top tripod is parallel to any in the bottom, inverted tripod.

It is possible that an image involving the hexahedron schema only represents the orientations of edges with respect to the top/bottom direction. This

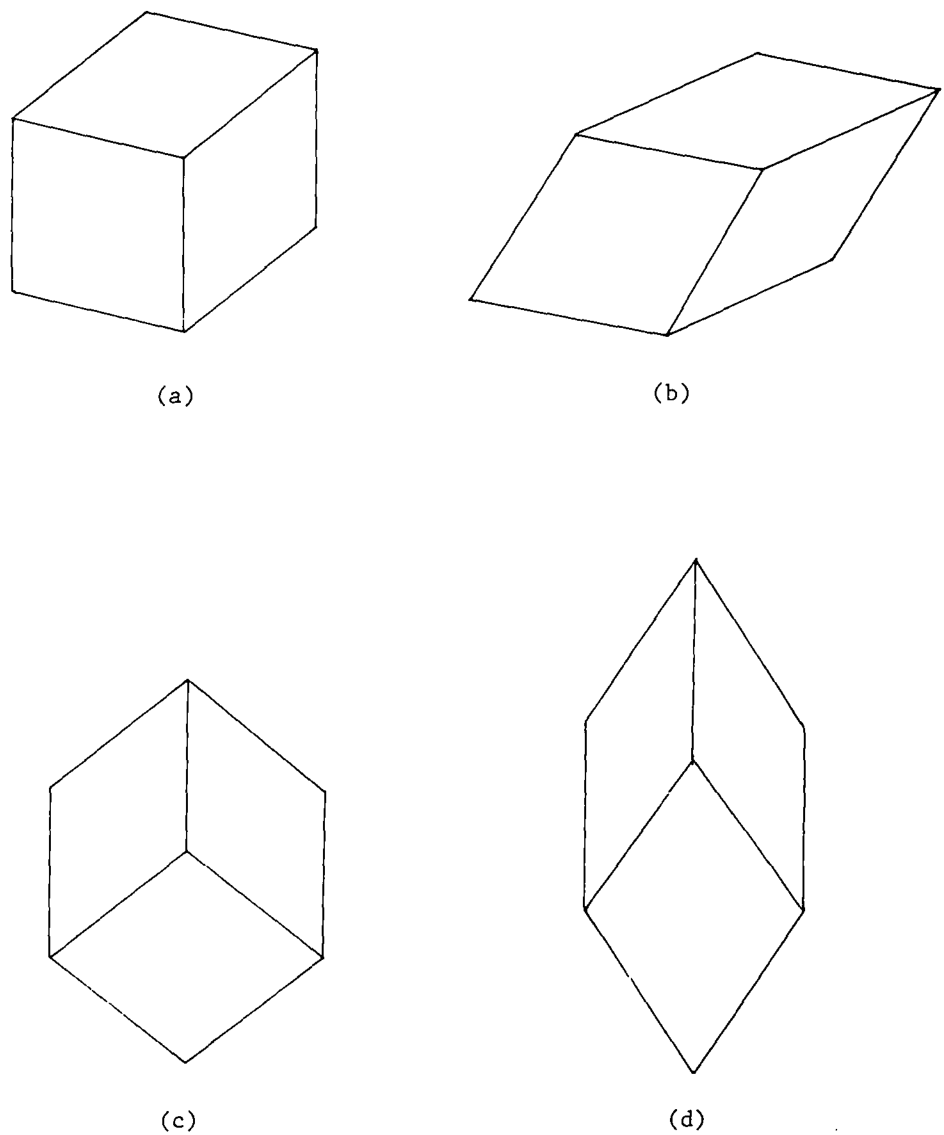


Figure 4 A perfect cube (a) and a distorted version of it (b). When the same two figures are perceived as having an intrinsic top/bottom direction aligned with their axis of threefold rotational symmetry, the cube (c) does not appear more regular than the noncube (d). The effect is more powerful when the real objects are observed as they are deformed.

would explain why judgments about parallelism are difficult. However, there also seem to be other relevant facts which are clearer if we just consider the central zigzag ring of a hexahedron. Figure 5 shows this ring embedded in a context which encourages a particular set of orthogonal directions to be assigned to the ring. It is possible to see the ring as consisting of a tilted central rectangle with triangular flaps folded in opposite directions, one on each side. When a mental image of this structure is formed, it is obvious that the edges forming the two ends of the rectangle are parallel, but not obvious that there are any other parallel pairs of edges.

It seems that it is hard to compare the orientations of edges which belong to different substructures of a mental image (the triangular flaps) especially if the edges do not align with the direction systems used for the substructures, and these systems in turn are not aligned with each other. There are many possible reasons for this difficulty, but whatever the reason, the effect is that the particular structural description used determines which judgments about parallelism are hard and which are easy. The zigzag ring can be parsed into a central tilted

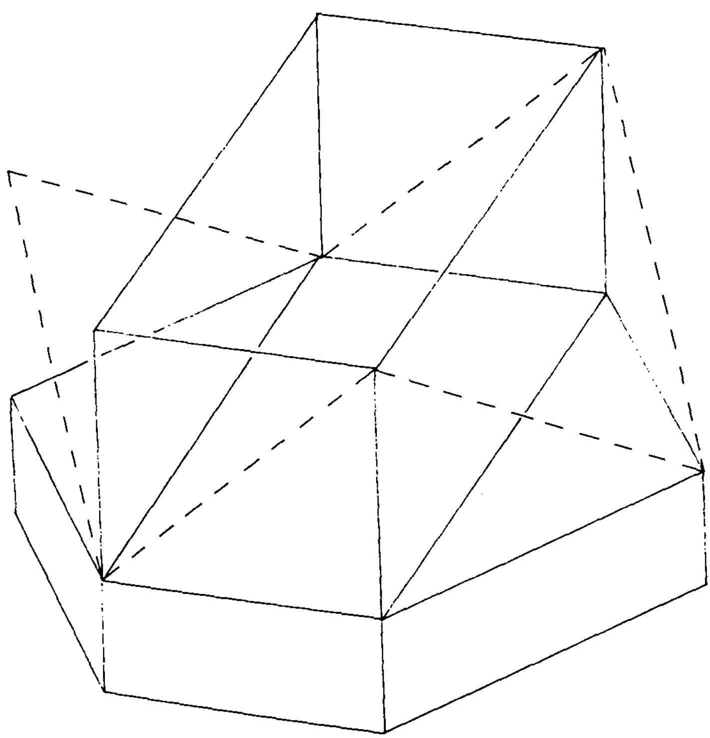


Figure 5 A wire-frame cuboid and a solid hexagonal prism provide a context which helps the central ring of a hexahedron (dashed lines) to be seen as a central tilted rectangle with oppositely tilted triangular flaps on each side.

rectangle and two triangular flaps in three different ways, each of which makes apparent a different parallel relationship.

7. REPARSING MENTAL IMAGES

A few subjects discovered a strategy which enabled them to form an approximately correct mental image of a cube standing on one corner, as required in task 1. The following instructions may enable the reader to duplicate the strategy. First, re-imagine the cube as two horizontal diamonds, one above the other, connected by four verticals. Imagine that one of the verticals is directly in front so that the axes of the diamonds point toward you. Notice that between them the diamonds contain all the corners. Now tip this diamond prism backward with the far corner of the bottom diamond remaining on the table. Notice that by tipping the prism by the correct amount, it must be possible to position the front corner of the top diamond vertically above the back corner of the bottom one, though it may be hard to imagine this relationship while keeping a clear image of the diamonds.

Some people can use a mental image constructed in this way to point out the corners of the tilted cube. They use their stationary fingertip as an anchor point for the top diamond, and indicate where its remaining three corners come. Then they use the place on the tabletop directly below their stationary fingertip to anchor the bottom diamond whose remaining three corners can be indicated. An interesting deficiency of this strategy is that subjects are unable to describe the relationships between the six corners they have pointed out. They can see them as forming two groups of three, but they cannot see them as forming a zigzag ring, nor do they realize that one corner from each group is at the same height as two from the other. They seem unable to notice relationships which are not easily expressed in terms of the particular groupings and direction systems that they are using for their mental image. As a result of this inability, perhaps, they are unable to reparse their mental image into different groups with different intrinsic direction systems, which would make these relationships evident.

There are many other cases of mental images which are hard to reparse. For example, the central zigzag ring of a hexahedron can be seen as a kind of crown consisting of three triangular flaps folded upward and outward from a horizontal triangular base (see figure 6). Given this mental image, it is very hard to realize, without using perception, that by grouping the edges differently the same structure can also be represented as an upside-down crown, or as a central tilted rectangle with triangular flaps on each side as shown in figure 5.

8. THE ANALOG COMPONENT OF MENTAL IMAGES

The great difficulty we have in reparsing mental images is evidence against the view that images are "analogical" representations of the world that are quite separate from the structural descriptions we assign. If this were so, it should be

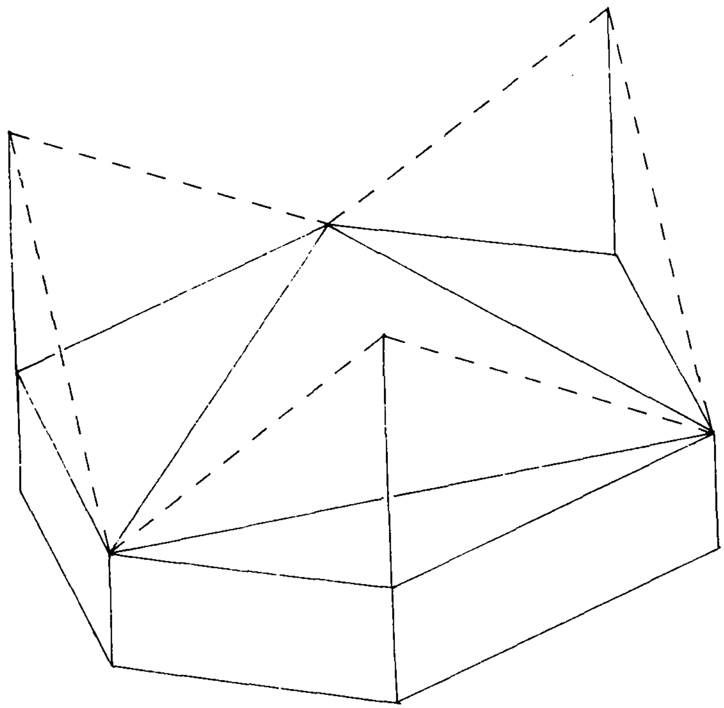


Figure 6 A solid hexagonal prism with three vertical spikes which help the central ring of a hexahedron (dashed lines) to be seen as a "crown" with three triangular flaps sloping upward and outward from a triangular base.

easy to assign a different structural description to an image, just as it is easy to reparse a visible object.

The opposing view is that images consist of structural descriptions, or at least that such descriptions are an essential constituent of images. This view has been contested on the grounds that it cannot explain analog transformations like mental rotation (Paivio, 1977, p. 64). The rest of this section is therefore devoted to showing that such transformations are explicable in terms of structural descriptions.

There has been a tendency within artifical intelligence (Hollerbach, 1975; Minsky, 1975) to handle real-valued variables by subdividing their ranges into a small number of intervals and using a discrete label to represent all values within a given interval. For example, the relational networks that are used to represent spatial structures often have arcs with labels like "above" or "left-of" (Winston, 1975). In some cases, like the convex/concave labels of Huffman

(1971) and Clowes (1971), the particular quantization used can be justified by showing that qualitatively different inferences follow from the different categories. In other cases, the use of discrete labels may have allowed programmers to concentrate on other more interesting aspects of their theories without worrying about the representation of real-valued variables. However, there is nothing to prevent real-valued labels from being attached to structural descriptions. Indeed, a very natural way to represent the orientation of an edge is to use the angles it makes with the assigned directions of the structure containing it (Anderson, 1978). Similarly, lengths can be represented by real-valued labels once a scale has been chosen. Given numerical labels, continuous changes like rotation or dilation can be simulated by continuously changing the numbers without altering the actual structure of the representation.

There are several phenomena which support the idea that mental rotation involves changing real-valued labels on a structural description. Metzler (1973) found that mental rotations are easier if they are about an axis which aligns with an intrinsic direction of the object. During a real rotation, the relationship of the object to its context changes, so during a mental rotation it is necessary to change the labels which indicate how the object's intrinsic directions are oriented with respect to the contextual direction system. If the relationship between one of these intrinsic directions and the context is unaffected by the rotation, less labels need to be changed and the mental rotation should be easier.

Mental folding (e.g., imagining a pen-knife opening and closing), provides further evidence that analog transformations involve structural descriptions. Shepard and Feng (1972) have shown, using reaction times, that mental folding involves internal representations which change either continuously or in rather small steps. An important aspect of mental folding is that for a given structure, the way it is parsed determines which folds can be imagined. Imagine the central zigzag ring of a hexahedron using the ''crown'' schema in which the edges form three triangular flaps pointing upward and outward from a horizontal triangular base (see figure 6). Now imagine the flaps folding downward through the horizontal plane until they form an upside-down crown. Many subjects report that they can do this easily, whereas it is extremely difficult to imagine the same physical transformation if one uses the schema in which the ring is composed of a central tilted rectangle with a triangular flap on each long side as in figure 5.

Using the crown schema the required transformation can be imagined by changing the labels which indicate the orientations of the three triangular flaps with respect to the whole object. The threefold rotational symmetry of the crown would allow the three flaps to share an orientation label (see figure 3b), so it is possible that only one label needs to be changed. With the other schema, however, the same physical transformation involves changing the shapes of two of the parts as well as changing the orientations of all three. So as well as changing the labels which represent the relationship of the parts to the whole, it is necessary to change some of the labels that relate the individual edges to the parts. This may even involve minor changes to the structural description, since the ends of the central rectangle do not remain parallel, and rectangles may only have one numerical label to represent the orientation of both ends. Mental folding, like mental rotation, shows that the difficulty in imagining a continuous physical transformation depends on how many continuous variables need to be changed in the particular structural description being used. It is sometimes argued that structural description theories cannot explain the continuity of mental transformations. Structural descriptions, it is claimed, do not need to pass through intermediate states, because it is as easy to compute large changes as small ones. This argument presupposes that if there are any numerical values in a structural description, they are represented and manipulated in the brain in much the same way as in a digital computer. An alternative model is the analog computer in which constraints between variables are implemented by feedback loops, and the values of variables have to change continuously. If the real-valued variables in a structural description are implemented this way, then mental transformations that involve changing variable values have to be continuous.

9. SUMMARY

Mental imagery tasks and perceptual demonstrations have been used to illustrate the involvement of structural descriptions in the imagination and perception of 3-D spatial structures. It has been shown that it is very difficult to imagine a cube when its orientation is specified in a way that is hard to relate to its usual structural description. An important technique has been to choose objects which have alternative structural descriptions and to observe the differential effects of those descriptions on subjects' abilities to create and transform mental images, or to notice relationships within them. Since the imagined external objects are the same, the differences cannot be explained in terms of unsegmented, viewer-centered representations.

The difficulty in changing structural descriptions for an imagined object has been illustrated and used to argue that such descriptions must form an integral part of a mental image. This claim has been further corroborated by demonstrations showing that the difficulty of analog transformations like mental folding is strongly affected by the choice between alternative structural descriptions. The description in which the fewest continuous variables need to be changed is the one that makes the transformation easiest to imagine.

As well as showing the involvement of structural descriptions in mental imagery, the demonstrations have provided evidence that those descriptions are hierarchically organized, that they have assigned significant directions at each level, and that they use continuous variables for representing metric information.

ACKNOWLEDGMENTS

I wish to thank Max Clowes, Steve Draper, Phil Johnson-Laird, Don Norman, Dave Rumelhart, and Aaron Sloman for helpful comments on the manuscript.

REFERENCES

Anderson, J. R. Arguments concerning representations for mental imagery. *Psychological Review*, 1978, *85*, 249–277.

Attneave, F. Triangles as ambiguous figures. *American Journal of Psychology*, 1968, *81*, 447–453.

Baylor, G. W. Program and protocol analysis on a mental imagery task. *Proceedings of the Second International Joint Conference on Artificial Intelligence*, 1971.

Clowes, M. B. Pictorial relationships—a syntactic approach. In Meltzer & Michie (Eds.), *Machine intelligence 4*. Edinburgh: Edinburgh University Press, 1969.

Clowes, M. B. On seeing things. *Artificial Intelligence*, 1971, *2*, 79–112.

Cooper, L. A. Demonstration of a mental analog of external rotation. *Perception and Psychophysics*, 1976, *19*, 296–302.

Cooper, L. A. & Shepard, R. N. Chronometric studies in the rotation of mental images. In W. G. Chase (Eds.), *Visual information processing*. New York: Academic Press, 1973.

Gregory, R. L. *Eye and brain*. London: Weidenfeld & Nicholson, 1966.

Hollerbach, J. M. Hierarchical shape description of objects by selection and modification of prototypes. *AI-TR-346*, M.I.T. AI Laboratory, Cambridge, Mass., 1975.

Huffman, D. A. Impossible objects as nonsense sentences. In Meltzer & Michie (Eds.), *Machine intelligence 4*. Edinburgh: Edinburgh University Press, 1971.

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

Kuipers, B. J. Modeling spatial knowledge. *Cognitive Science*, 1978, *2*, 129–153.

Leeuwenberg, E. A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, 1971, *84*, 307–349.

Marr, D. Representing visual information. In A. R. Hanson & E. M. Riseman (Eds.), *Computer vision systems*. New York: Academic Press, 1978.

Marr, D. & Nishihara, H. K. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, Series B*, 1978, *200*, 269–294.

Metzler, J. Cognitive analogues of the rotation of three-dimensional objects. PhD. dissertation, Stanford University, 1973.

Miller, G. A. & Johnson-Laird, P. N. *Language and perception*. Cambridge University Press, 1976.

Minsky, M. L. A framework for presenting knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Minsky, M. L. A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Narasimhan, R. Syntax directed interpretation of classes of pictures. *Association for Computing Machinery*, 1966, *9*, 166–173.

Paivio, A. Images, propositions and knowledge. In J. M. Nicholas (Eds.), *Images, perception and knowledge*. Dordrecht, Netherlands: Reidel, 1977.

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975.

Palmer, S. E. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977, *9*, 441–474.

Perkins, D. N. How good a bet is good form? *Perception*, 1976, *5*, 393–406.

Podgorny, P. & Shepard, R. N. Functional representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 21–35.

Pylyshyn, Z. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, *80*, 1–24.

Reed, S. K. Structural descriptions and the limitations of visual images. *Memory and Cognition*, 1974, *2*, 329–336.

Rock, I. *Orientation and form*. New York: Academic Press, 1973.

Shepard, R. N. & Feng, C. A chronometric study of mental paper folding. *Cognitive Psychology*, 1972, *3*, 228–243.

Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects. *Science*, 1971, *171*, 701–703.

Shneier, M. O. Object representation and recognition in machine vision. PhD. Thesis, University of Edinburgh, 1978.

Turner, K. J. Computer perception of curved objects using a television camera. PhD. Thesis, University of Edinburgh, 1974.

Ullman, S. The interpretation of visual motion. PhD. Thesis, M.I.T. AI Laboratory, Cambridge, Mass., 1977.

Winston, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.