In J. Long & A. Baddeley (Eds.)

Attention and Performance IX

Hillsdale N.J.: Lawrence Erlbaum, 1981

15

Frames of reference and mental imagery

Geoffrey E. Hinton
and
Lawrence M. Parsons
Center for Human Information Processing
University of California, San Diego, California
U.S.A.

ABSTRACT

Successively perceived parts of a scene or object must be related to one another to create a representation of the whole (Hochberg, 1968), and the same object when seen from a different viewpoint must be seen to have the same three-dimensional spatial structure. These perceptual requirements can be met by a system that explicitly represents and manipulates relationships between a viewer-centered frame of reference and frames of reference that are embedded in external objects. The computational apparatus that is required for handling these relationships during normal perception can also be used for simulating continuous spatial transformations and for performing imagery tasks. The use of a three-dimensional, viewer-centered frame of reference as a common space in which to coordinate the various frames embedded in objects may be what distinguishes visual imagery from other methods of spatial reasoning.

A major objection to the view that people explicitly represent and manipulate relationships between frames of reference is a finding by Cooper and Shepard (1973) that suggests that people cannot mentally rotate an abstract frame of reference. We present an experiment that shows that under suitable conditions, predicted by our theory, people do mentally rotate an abstract frame.

A THEORY OF SPATIAL REPRESENTATIONS

Figure 15.1 depicts a three-dimensional spatial structure composed of six rods. To help depict the structure it is embedded in a solid object. When people are shown the configuration of six rods by itself, they can perceive it in several distinct ways. They can "parse" it into the groups *ab*, *cd*, *ef*, and see it as a "crown" consisting of three triangular flaps that slope upward and outward. Alternatively, they can group *a* and *d* together as the ends of a central tilted rectangle and see *b* and *c* as a triangular flap sloping downward and outward and *e* and *f* as a flap sloping upward. Figure 15.2 shows two structural descriptions that represent these alternative perceptions.

Each part of the structure appears to have its own intrinsic frame of reference. One of the triangular flaps in the crown interpretation, for example, has an intrinsic frame that has three orthogonal directions defined by the axis of symmetry of the triangle, the normal to the plane of the triangle, and the "sideways" direction parallel to the base of the triangle. Each flap defines a different intrinsic frame of reference, but relative to its own frame of reference each flap is identical. Rock (1973) provides considerable experimental evidence for intrinsic frames of reference, and Marr and Nishihara (1978) and Hinton (1979a) discuss their theoretical importance.

An intrinsic frame of reference serves two functions. The object can be given a shape description that is independent of the subject's viewpoint by describing the orientations and dispositions of the object's features relative to its own intrinsic frame. That is why the three flaps in the crown are seen as having the same shape even though their retinal images are very different. Second, the relationship of an object to its context can be specified by specifying the relation-

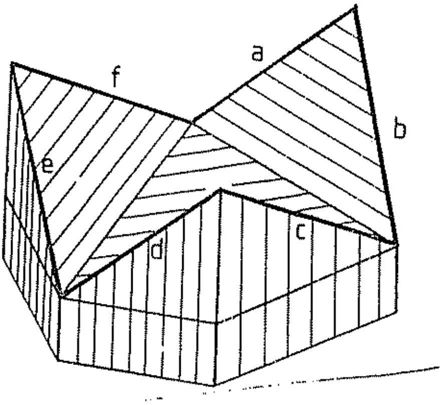

FIG. 15.1. The spatial structure formed by the six rods (heavy lines) can be perceived either as a "crown" composed of three triangular flaps or as a "zigzag" composed of a central rectangle with one triangular flap sloping up from the bottom edge and another sloping down from the top edge.

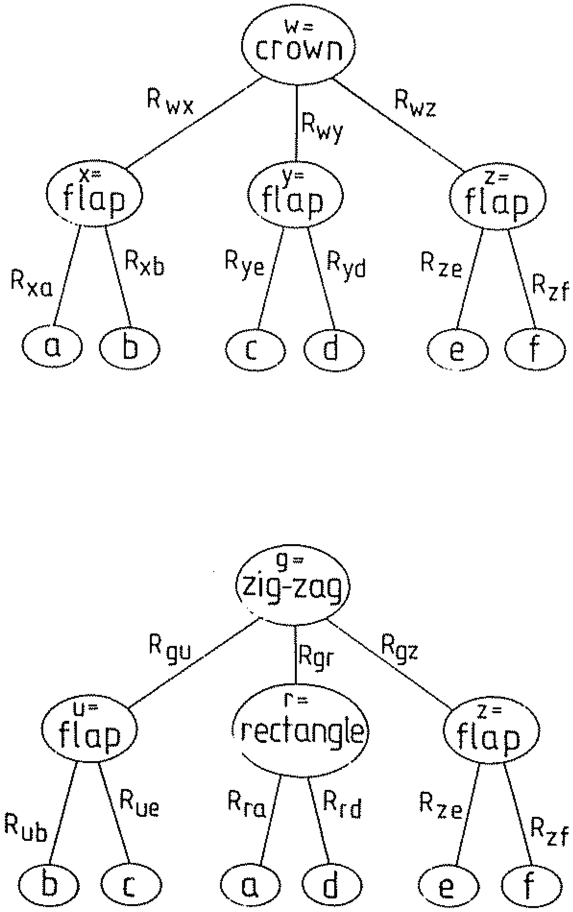

FIG. 15.2. Two alternative structural descriptions corresponding to the alternative ways of seeing the six rods in Fig. 15.1. The nodes represent objects or their parts, and the labels on the arcs represent spatial relationships.

ship between the object's intrinsic frame of reference and an intrinsic frame of reference embedded within the context. The relationship of one flap to the whole crown, for example, is determined by the relationship of the intrinsic frame of the flap to the intrinsic frame of reference of the whole crown.

Relationships between Frames of Reference

Frames of reference themselves cannot be directly described. They can only be specified by their relationships to other frames of reference. If nonrigid transformations like shear and elongation are discounted, the relationship between two

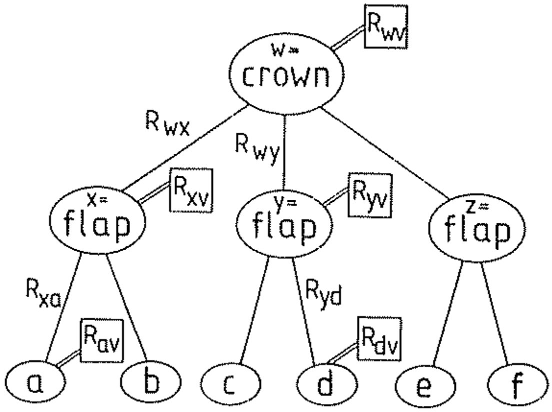

FIG. 15.3 A structural description with some attached viewpoint information in the form of relationships like R_{av} that specifies how the intrinsic frame of a is related to the common, viewer-centered frame.

three-dimensional frames of reference has 7 degrees of freedom: 3 for translation, 3 for rotation, and 1 for scale. So relationships between two intrinsic frames of reference could be represented by 7 real-valued variables. It is unlikely that the representation of a spatial relationship in the brain is anything like the natural representation in a digital computer, but whatever it is like, it must have the same number of degrees of freedom each of which must be capable of having a more or less continuous range of values.

It is possible to simulate continuous spatial transformations (Shepard & Feng, 1972; Shepard & Metzler, 1971) by simply modifying the representations of the spatial relationships in a structural description. For example, the transformation in which the three triangular flaps of the crown fold outward and downward can be simulated by changing R_{wx}, R_{wy}, and R_{wz} in Fig. 15.2. There is evidence that this is what is happening when people imagine a continuous transformation (Hinton, 1979a). Each structural description makes some relationships explicit and leaves others implicit, so the number of explicit relationships that have to be changed to simulate a given physical transformation may depend on which of several alternative structural descriptions is used. People find transformations easier to imagine if they use a structural description in which only a few of the explicit relationships need to be changed.

The structural descriptions shown in Fig. 15.2 are useful for recognizing objects and providing stable representations of the structure of the environment because they are independent of viewpoint. However, in addition to this kind of viewpoint-independent knowledge, people are also aware of how objects and their parts are related to the current viewer-centered frame of reference. Figure 15.3 shows how this knowledge can be represented by attaching to all or some of the object nodes in a structural description information about the relationship between the intrinsic frame of reference of that object and the current viewer-centered frame.

Inferring Relationships between Frames of Reference

Some of the relationships between frames can be computed from others, and a visual system that uses intrinsic frames needs to be able to do this. One type of computation yields the intrinsic relationship between two intrinsic frames, *a*, *b*, from information about the relationship of each frame to a common, viewer-centered frame, *v*. This kind of computation must occur when we "just see" an intrinsic relationship because what is perceptually available is the relationship of each object to the viewer-centered frame. The computation may be expressed as follows:

$$R_{av} \ \& \ R_{bv} \Rightarrow R_{ab}$$

Another type of computation must be performed when an object has been perceived and recognized as a whole, and the system uses its stored knowledge of the spatial structure of the object to help it pick out a particular part of the object. The system must figure out where the part is in viewer-centered terms so that it can make the appropriate eye movement or internal change of attention. So the relationship of the part to the viewer-centered frame must be computed from the intrinsic relationship between the whole, *a*, and the part, *c*, and the relationship of the whole to the viewer-centered frame, *v*:

$$R_{av} \ \& \ R_{ac} \Rightarrow R_{cv}$$

If these two types of computation are occuring all the time during normal perception, it is reasonable to suppose that people have special-purpose hardware for implementing them efficiently. We show that this same hardware could also be used in performing imagery tasks.

Mental Images and Pictures

When a person forms a mental image of a spatial structure, he/she can often "just see" new spatial relationships that were not explicitly used in forming the image. Imagine, for example, the following journey. Go a mile north, then a mile east, and then a mile north again. Now, what is the direction back to your starting point? Most people report that they form an image and read off the answer without any conscious inference. It seems to them that they create something like a picture in their mind that they can then perceive. A crucial question is: In what ways do the internal representations involved in imagery resemble pictures and in what ways to they differ?

There are two separable properties of pictorial or arraylike representations, and we shall argue that mental images have one of these properties but not the other.

To create a picture of an object or scene it is necessary to adopt a specific viewpoint. This is equivalent to choosing a relationship between a frame of reference embedded in the scene and the frame of reference of the picture. As well as this "commitment to viewpoint," pictures have a further property that we shall call "atomic depiction". For simplicity we shall assume that a picture is like an array in which each cell may be given a number of properties like color or intensity but cannot have internal spatial structure. Notice that it is the mapping from scene to picture and the structure of the pictorial medium, not the structure of the scene itself, that defines how the scene is carved into separate atomic parts each of which is depicted by averaging its visible properties.

A representation can have the property of commitment to viewpoint without necessarily having the further property of atomic depiction. The kind of representation in which each node in a structural description is given a relationship to a common, viewer-centered frame of reference is committed to a specific viewpoint, but it does not require decomposition into elements defined by the grain size of the pictorial medium. Instead, the viewpoint information is attached to the units that are meaningful within the scene, and it is possible to attach viewpoint information to some units without necessarily attaching it to all. For example, it is possible to represent where a large elephant is in viewer-centered space without being forced to represent the orientation of its trunk.

Much of the evidence that is normally taken to corroborate the idea that a mental image is like a picture actually only shows that mental images are committed to viewpoint. One line of evidence for the further property of atomic depiction is that people seem to zoom in mentally in order to see details in mental images. Kosslyn (1975) claims that zooming is necessary because mental images have a finite grain size. However, Hinton (1979b) shows how the need for zooming is also predicted by a model in which there is some noise in the representation of the parameters of a spatial relationship.

The following section shows why commitment to viewpoint is a computationally useful property in solving mental imagery tasks.

Computing Spatial Relationships in Mental Images

It is possible to define a spatial structure by giving some but not all of the spatial relationships between its parts. The remaining relationships can then be inferred from the ones that are explicitly given. The imagery task presented in the previous section required this kind of inference. The obvious way of inferring the relationships between two nodes in a structural description that are not directly connected is to find an indirect pathway of known intrinsic relationships. Each relationship is equivalent to a matrix, and the product of these matrices is the required relationship between the two nodes.

There is, however, an alternative method of computation that requires commitment to viewpoint and is therefore a more plausible model of what occurs during visual imagery. The computation involves three stages:

1. One of the nodes in the structural description is given a relationship to the viewer-centered frame of reference. This relationship can be chosen arbitrarily. In effect, it defines the imagined viewpoint.
2. Consistent relationships to the viewer-centered frame can then be propagated to other nodes by using the intrinsic relationships and computations of the form: R_{av} & $R_{ab} \Rightarrow R_{bv}$. The process of propagating consistent relationships is what is required to form a mental image.
3. Finally, *any* relationship between two nodes can be computed immediately by using a computation of the form: R_{av} & $R_{bv} \Rightarrow R_{ab}$. Notice that this is the same primitive computation as is used during perception to "just see" a relationship. This may explain why people introspectively describe the process of computing a new relationship in a mental image as "just seeing" it.

This method enables a system that uses a hierarchy of intrinsic frames to compute an unperceived relationship by making use of the mechanisms that must already be available for normal perception. It is like the process of drawing a picture in that it uses the intrinsic relationships to give every object a consistent relationship to a common, viewer-centered frame. Once everything has been related to one frame, it is possible to read off implicit relationships without performing long chains of inference. However, the computational advantages of a common frame of reference are achieved without requiring atomic depiction. Objects do not need to be decomposed into pieces of a size defined by the picture grain or the individual array cells in order to represent their relationships to the common frame of reference.

THE FUNCTION OF MENTAL ROTATION

A major problem for the theory just presented is why mental rotation is necessary at all for the kind of task in which a subject has to judge whether two objects at different orientations are the same or are mirror images of each other. Shepard (1979), Kosslyn, Pinker, Smith, and Shwartz (1979), and Pinker and Finke (1980) argue that the fact that people use mental rotation is evidence that they do not have a representation of the shape of an object that is independent of the object's orientation. If they did, the argument goes, they should be able to judge the identity or nonidentity of two shapes in different orientations without performing mental rotation.

This argument certainly appears to rule out any theory that claims that the internal representation of the shape of an object is generated by imposing an intrinsic frame of reference on the object and describing its features relative to

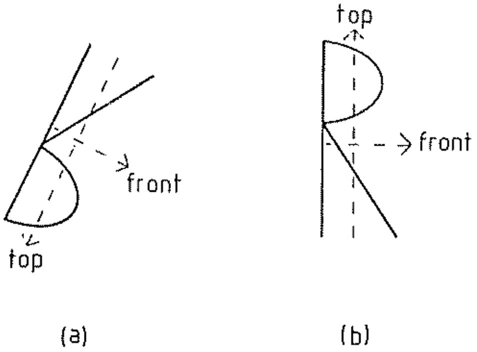

FIG. 15.4. Two versions of an R, with the intrinsic frames of reference that people impose on them. Relative to their intrinsic frames, they have identical features. The differences lie in the orientation, position, and handedness of the imposed frames.

that frame of reference. Given two objects in different orientations, it should be possible to impose appropriately oriented intrinsic frames on each of them and then to check whether the resulting shape descriptions were identical. It is hard to see how this kind of process could account for the mental rotation data. The time take to assign tilted intrinsic frames might depend on the tilt, but this would make the wrong predictions. In experiments where two objects are presented simultaneously, it would predict that the reaction time should depend on the sum of the two tilts, not on the difference. Furthermore, it would not explain subjects' introspections that they mentally rotate one of the objects from its current orientation into alignment with the other object.

There is, however, a simple additional assumption that makes the theory being proposed entirely consistent with the mental rotation data. Mental rotation, it will be argued, is necessary to overcome a peculiar and normally irrelevant deficit in our way of representing spatial structures: Although we can rapidly perceive the shapes of objects in unusual orientations, we do not normally know and are not normally interested in the handedness of these objects. (*Handedness* is the property that distinguishes a right-hand glove from a left-hand one.)

Introspectively, we do not perform mental rotation in order to recognize a tilted object like the *R* in Fig. 15.4a. We can see that it is an *R* and that it is roughly upside down without any mental rotation. Indeed, we must be able to identify the letter and to see which way up it is in order to decide what rotation to perform to make it upright. Cavanagh (1977) has shown that the effects of orientation on the time required to identify letters are much smaller than the times required for mental rotation.

Mental rotation does seem to be required in order to decide whether the *R* is a normal or mirror-image version (i.e., to decide on its handedness). It appears that we can only compare the handedness of two objects if their orientations are aligned, and we use mental rotation to achieve this alignment. Deciding whether the *R* is forward or backward is a special case in which one of the objects is remembered rather than perceived.

The need for alignment in judging relative handedness is a subtle consequence of the assignment of intrinsic directions to an object. The problem is that we can assign either a left-handed or a right-handed intrinsic frame, and we use the intrinsic frame that yields a familiar shape representation (see Fig. 15.4a). If there are two objects of opposite handedness, we use intrinsic frames of opposite handedness and obtain the same shape representation. This means that to compare the handedness of the objects it is necessary to compare the handedness of their intrinsic frames. This would be possible without any mental rotation if we knew the absolute handedness of each frame. It appears, however, that we do not have an explicit representation of the handedness of an intrinsic frame (this is our additional assumption). So, in order to compare handedness we mentally rotate one frame until all but one of its significant directions align with the corresponding directions in the other frame, and then we compare the remaining direction. In two dimensions this means rotating one frame until its top/bottom direction aligns with the other one and then comparing the front/back directions. (Generally, equivalent arguments about handedness apply in both two dimensions and three dimensions, but they are easier to present in two dimensions.)

Mental rotation, in our theory, does not affect the representation of the object's shape. It simply involves altering the explicit representations of two spatial relationships, one between the object and the viewer and the other between the object and its context. The alterations must preserve handedness (a property of continuous rotations), but this requirement does not, in itself, explain why mental rotation appears to be continuous. There are many possible reasons. If, for example, people only have the computational hardware for rotating through a small angle, then large rotations would require repeated use of this hardware, just as large shifts require repeated operations in a simple shift register. We must emphasize, however, that the aim of this chapter is not to explain why mental rotation times depend on angle of rotation. We merely aim to show that the phenomena of mental rotation are compatible with the assignment of intrinsic frames of reference for generating shape descriptions.

The idea that we can know both the top/bottom direction and the front/back direction of a two-dimensional intrinsic frame and yet not know its handedness may appear strange, because these two directions determine the handedness of the frame. However, neither direction by itself determines handedness, so a system that represents the two directions separately may lack an explicit representation of the handedness, even though this is implicit in the representations of the two directions.

It is not theoretically necessary to use a continuous rotation to bring two frames into alignment. A sequence of more discrete operations could also be used. For example, the top/bottom direction could be changed by 30° and the front/back direction could then be changed by the same amount so as to restore perpendicularity of the two directions. It is not possible, however, to use large steps like 90°, because it would then be possible to reverse the handedness of the frame when restoring perpendicularity. Reversal could not occur if the top/bottom and front/back directions were changed either both clockwise or both counterclockwise. However, this strategy involves labeling angles as clockwise or counterclockwise, which is equivalent to knowing the handedness of systems in arbitrary orientations. It is just this kind of explicit representation of handedness that we are postulating is absent in people.

Normal adults must have knowledge that is equivalent to knowing the absolute handedness of a vertically aligned frame of reference because they can distinguish correct letters from backward ones. However, this knowledge may consist in knowing which way the front of a normal version points when the character is upright. This would mean that knowledge of shape and handedness were separate, even for characters whose handedness we know. Some such separation appears to be necessary to explian how certain dyslexics can know the shapes of characters but not know which way round they go. Explicit knowledge of handedness is normally ecologically irrelevant. There are very few objects in the world whose properties depend on their handedness, and for man-made artifacts like writing where handedness is crucial, people have unusual difficulty. Normally, we want to classify objects that differ only in handedness as having the same shape. A person's profile seen from the other side is a good example.

In order to reconcile the need for mental rotation with the idea that people achieve invariant representations of the shapes of objects by assigning intrinsic frames of reference, we have postulated that people do not know the handedness of the tilted frames that they assign.

THE EXPERIMENT

Introduction

The hypothesis about the function of mental rotation predicts that it should be possible to rotate mentally an upright frame of reference of known handedness to the appropriate orientation before a tilted letter is presented. The subject could then judge whether the letter was forward or backward without performing any further mental rotation. Cooper and Shepard (1973) claim that people cannot do this under their particular experimental conditions. They found that giving subjects advance information about the orientation of a letter did not remove the need to rotate the letter mentally to upright in order to judge its handedness. This is an important result because it suggests that people cannot use the advance orientation information to rotate an abstract frame of reference of known handedness to the appropriate orientation, and this inability corroborates the view that shape and orientation are not separately represented.

We decided to investigate whether Cooper and Shepard's (1973) results might depend on special characteristics of their experimental design. In particular, we were interested in whether the subjects had simply failed to discover the strategy of rotating a frame of reference and whether the difficulty of using this alternative strategy might depend on the particular character set being used and the particular form of the required response.

Cooper and Shepard (1973) presented subjects with a character from the set G, J, R, 2, 5, 7, displayed in one of a number of orientations, and required them to respond with a right-hand button press if the character was a normal version and a left-hand button press if it was a mirror image. We reasoned that the particular character set might make it hard to put advance information about orientation to good use because there is no common relationship to a frame of reference that is shared by just the normal versions. The characters in the set F, R, G, L, on the other hand, all seem to have a "front" that faces to the right, so the normal versions all "agree" with a frame of reference that has its front on its right. This allows subjects to use the following strategy: When they see the arrow indicating the top/bottom direction of the upcoming character, they mentally rotate an abstract frame of reference that initially has a vertically aligned intrinsic top/bottom direction and a front pointing to the right. They note which way the *front* of this frame points when its top/bottom direction aligns with the arrow. When the character appears, they have only to judge whether its front points in the same or the opposite direction.

In a pilot experiment, we discovered that when a normal character was upside down so that its front was on the left of the screen, there was a strong tendency to press the left-hand button, because the button location "agreed" with the direction in which the front of the character pointed. To avoid this type of intrusion of the spatial characteristics of the response, we required subjects to press a key if the character was normal and to make no response to mirror images.

One group of subjects was explicitly instructed in how to make use of the advance orientation information. They were given trials both with and without the advance information. Initially, they were given characters from the set F, R, G, L. Later, they were given characters from the set F, R, J, 7, and they were instructed to try to see the J and the 7 as pointing to the right.

A second group of subjects was used as a control to show that it was the particular character set and the explicit instruction in the strategy that caused us to obtain different results from Cooper and Shepard (1973). The group was presented with characters from the set F, R, J, 7, both with and without advance orientation information, but they were not explicitly instructed in the strategy.

Method

Subjects. Twelve right-handed University of California students participated in this study for credit in a lower-division course in psychology.

Stimuli. The test stimuli were asymmetrical alphanumeric characters—one Arabic numeral, 7, and five uppercase letters, F, R, G, L, J. Each of the six characters could be presented in any of 12 equally spaced orientations in the picture plane, in either its normal or mirror-image version. Advance information for the orientation of the upcoming character consisted of an arrow drawn through the center of the field and pointing to the position at which the top of the character was about to appear. All stimuli were displayed visually in the center of the bit-mapped display of a TERAK microcomputer. They subtended about 5° of visual angle.

Design. The 12 participants were randomly divided into 6 experimental and 6 control subjects. Experimental subjects performed 24 blocks (of 96 trials each) as follows. The first 12 blocks consisted of trials with the character set F, R, G, L. In half of these blocks, a trial was always preceded by advance information for orientation. In the other 6 blocks, subjects received no advance information and the trial began with the onset of the test character. In the second half of the experiment (the second 12 blocks) the character set was F, R, J, 7 (two characters from the first set and two new characters). Again, all the trials in 6 of these blocks were preceded by advance orientation information, and all trials in the other 6 blocks were without the advance information.

The control subjects were told to do the best they could to use the arrow in preparing for the trial, but they were not instructed in the strategy. Apart from this, the control subjects exactly replicated the *second* 12 blocks (i.e., those with F, R, J, 7).

The order of conditions for the 12 blocks with each character set was identical for all subjects: Blocks 1, 2, 5, 6, 11, 12 were in the advance orientation information condition, and Blocks 3, 4, 7, 8, 9, 10 were in the no-advance information condition. The first four blocks with each character set provided subjects with practice in the two conditions. Practice trials were identical in all respects to test trials. This order of conditions was designed to equate effects due to practice for blocks with and without advance information for orientation and to minimize the number of times a subject had to change between the two conditions.

Each block contained a trial for every possible combination of one of the four characters, one of the 12 orientations, and normal or mirror-image version. The trials were randomly ordered for each subject.

Procedure. Subjects were seated before a CRT screen with their right-hand index finger on a key of the microcomputer keyboard. They were told the task was to press the key as quickly as possible if the character was a normal version, regardless of the character's orientation. If it was a mirror-image version, they were told not to press the key. They did not make any large head movements.

On trials without advance information for orientation, a stimulus character would appear on the screen until a response was made or 1.5 sec had elapsed with no response. Trials with advance information for orientation began with the display of the arrow for 2 sec, followed by a stimulus character as before For all blocks, there was a 2-sec intertrial interval

Subjects were given immediate feedback information as to the accuracy and duration of their responses. At the occurrence of an incorrect response, the microcomputer made a readily audible buzzing sound. Subjects' reaction times (RTs) in milliseconds were displayed in a column along the left-hand margin of the CRT screen. The RT (± 1 msec) and accuracy of each response were recorded by the computer.

Results and Discussion

Figure 15.5 shows the mean RTs of correct positive responses, as a function of the orientation of the test stimulus, for experimental subjects with the F, R, G, L character set in the advance and no-advance information conditions. Clockwise and counterclockwise rotations of equal magnitude yielded comparable RTs; therefore they have been combined for all analyses.

For 5 of the 6 experimental subjects, and for the group as a whole, the advance orientation informtion condition yielded significantly lower values than the no-advance information condition for both the mean RTs averaged over

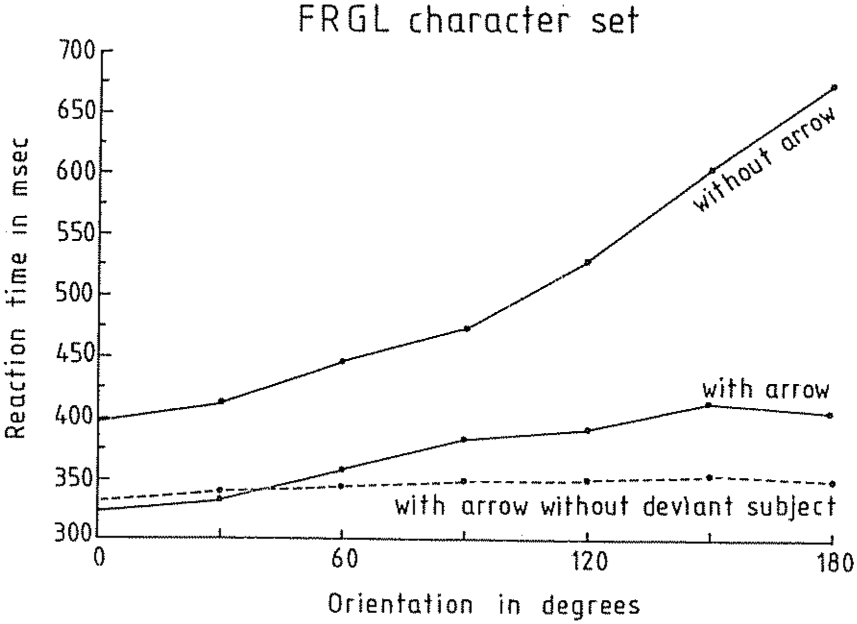

FIG. 15.5. Mean RT plotted as a function of orientation from vertical for the advance and no-advance information conditions The second plot of the advance information condition is without the data of the one subject who failed to follow the strategy that subjects were instructed to use

orientation [for the group, $t(19) = 3.34$, $p < .01$] and for the slope of the linear regression lines of RT on orientation [for the group, $t(925) = 3.15$, $p < .01$]. One subject appeared to be incapable of using the strategy. For the other 5 subjects, RT was only very slightly affected by orientation in the advance information conditions (see Fig. 15.5). The magnitude of this effect (19 msec in 180°) is much smaller than any reported times for mental rotation. In terms of our theory, this implies that it takes very slightly longer to impose a nonvertical frame of reference and to represent a stimulus relative to this frame.

The condition without advance information yields data typical of mental rotation. The linear regression equation accounts for 41% of the variance in the raw data, $t(1095) = 2.72$, $p < .01$. With advance orientation information, it accounts for 11%, $t(1098) = 2.81$, $p < .01$, but if the deviant subject is excluded, the variance accounted for drops to 1%. Even this is significant at the .01 level because of the large sample size. Subjects' mean error rates vary from 2% to 9%, with an overall mean of 6%. The error rates did not show any significant main effects or interactions of the factors of character, orientation, normal or mirror-image version, and advance or no-advance information.

As a result of the differences between our design and that used by Cooper and Shepard (1973), our subjects are able to use advance orientation information to prepare for the discrimination of a normal from a mirror-image character at various orientations. Moreover, subjects reported doing this by first rotating, as per instructions, a frame of reference that denotes top and front directions and then noting whether the front of the displayed character was facing in the direction specified by the frame of reference.

Our hypothesis was tested once more, this time using two members of the previous character set (F, R) and two new characters (J, 7). Figure 15.6 shows the correct positive RTs as a function of the angular difference from upright of the stimulus character. The figure shows the group data for both experimental and control subjects, with and without advance orientation information.

For control subjects, the linear regression equation accounts for a significant amount of the variance: In the advance informtion condition it accounts for 42% of the total variance, $t(1082) = 2.84$, $p < .01$, and in the no-advance information condition it accounts for 38% of the total variance, $t(1095) = 2.97$, $p < .01$. These subjects had significantly lower mean RTs for the advance than for the no-advance information condition (57 msec, $t(23) = 2.21$, $p < .05$). This difference may be due to a reduction in recognition time for the character when the top/bottom direction is known in advance or to the use of the advance orientation information to decide which way to rotate the character. For 4 of the 6 control subjects, and for the group as a whole, the advance orientation information did not significantly affect the slope of the regression line. For one of the two exceptional control subjects, both the regression slope and mean RTs were significantly lowered by advance orientation information, $t(185) = 2.21$, $p < .05$, and $t(358) = 2.02$, $p < .05$, respectively; conversely, for the other exceptional subject, advance orientation information significantly increased the slope of the

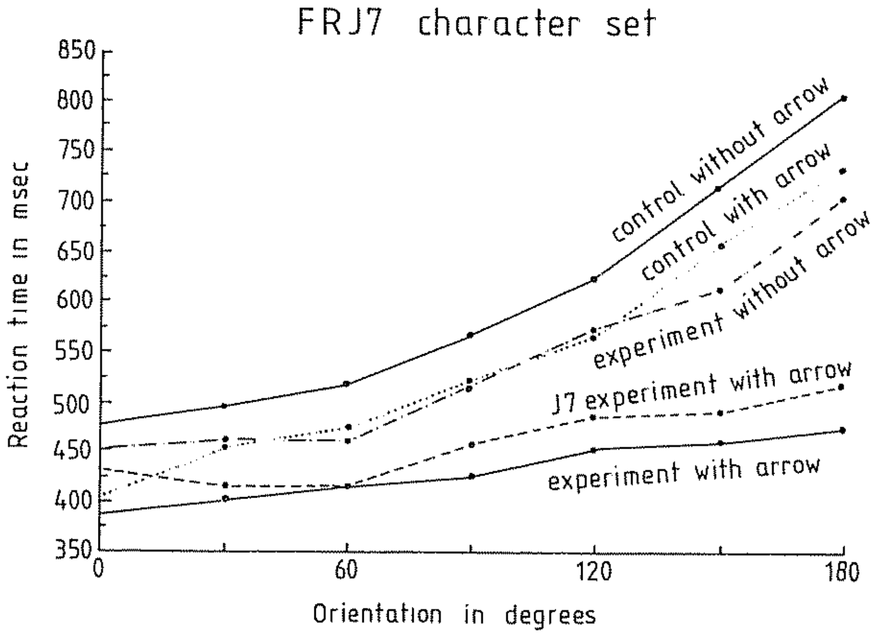

FIG. 15.6. Mean RT plotted as a function of orientation with and without advance information for experimental and control subjects. Data from the deviant experimental subject is not included. The data for the experimental subjects on the J and 7 trials is plotted separately to show that the advantage of advance information is not restricted to the F and R trials.

linear regression line, $t(188) = 2.15$, $p < .05$. The control subjects' overall error rates varied from 3% to 9% for individual subjects, with a mean of 5%. There were no significant interactions with any experimental factors.

Thus, the control subjects were apparently unable to use advance orientation information to avoid mental rotation in discriminating normal from mirror-image characters. This result, which can be considered a replication of the Cooper and Shepard (1973) finding, contrasts clearly with the experimental subjects' performance with this same character set (see Fig. 15.6). For 5 of these 6 experimental subjects, and for the group as a whole, there were significant effects of advance orientation information on the mean RTs averaged over orientation [for the group, $t(23) = 3.52$, $p < .01$]. (The subject failing to show this difference with the F, R, J, 7 character set was the one who also failed on the simpler F, R, G, L set.) Further, the RTs and regression-line slopes for just the trials with the two new characters J and 7 (see Fig. 15.6) again revealed significant differences between advance and no-advance information conditions [for 5 subjects, $t(19) = 3.01$, $p < .01$, and $t(450) = 2.97$, $p < .01$]. The small slope for the J and 7 trials with advance information (even omitting the deviant subject) suggests that some subjects occasionally revert to mental rotation for these characters.

The experimental subjects' overall error rates with the F, R, J, 7 set varied from 3% to 9% (with a mean of 6%). Again, the error data did not yield any significant effects

Thus, subjects with the appropriate instructions, training, and response conditions are able to use the advance orientation information to avoid mental rotation even with a character set in which there are no obvious features that discriminate normal characters from mirror images

SUMMARY

We have presented a specific alternative to the view that mental images are like pictures or arrays. The alternative is a hierarchical structural description containing explicit representations of quantitative spatial relationships. One type of relationship represents the intrinsic, viewpoint-independent spatial structure of a scene or object. A second type represents the relationships of objects and their parts to the viewer. These relationships to a viewer-centered frame of reference facilitate spatial reasoning, and it is their presence that characterizes mental imagery

Continuous spatial transforamtions like mental rotation can be mentally simulated by continuously changing the representations of the spatial relationships.

It is commonly claimed that models that include viewpoint-independent representations of shape cannot explain why mental rotation is necessary to compare two shapes at different orientations. We have shown how the normally useful ability to describe shapes relative to imposed frames of reference of either handedness may lead to a specific inability to compare the handedness of two nonaligned shapes or to judge whether a single nonupright shape has its normal handedness. Our argument requires us to postulate that people do not have an explicit representation of the handedness of the nonupright frames of reference that they impose on nonupright shapes.

Experimental evidence has been presented that shows that, with proper instruction and under conditions predicted by our theory, subjects can judge the handedness of a nonupright character without mentally rotating it. This result corroborates our explanation of why mental rotation is necessary and shows that its use is not incompatible with the existence of internal, viewpoint-independent representations of shape.

ACKNOWLEDGMENTS

This research was conducted while the first author was a Visiting Scholar with the Program in Cognitive Science, supported by the Sloan Foundation. The second author was supported by Grant No. BNS 7915336 from the National Science Foundation to George Mandler.

REFERENCES

Cavanagh, P. Locus of rotation effects in recognition. *Bulletin of the Psychonomic Society*, 1977, *10*, 152-154.

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

Hinton, G. E. Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science*, 1979, *3*, 231-250. (a)

Hinton, G. E. Imagery without arrays. Commentary on: S. M. Kosslyn, S. Pinker, G. E. Smith, and S. P. Shwartz, "On the demystification of mental imagery." *The Behavioral and Brain Sciences*, 1979, *2*, 555-556. (b)

Hochberg, J. In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception*. New York: Holt, Rinehart and Winston, 1968.

Kosslyn, S. M. Information representation in visual figures. *Cognitive Psychology*, 1975, *7*, 341-370.

Kosslyn, S. M., Pinker, S., Smith, G. E., & Shwartz, S. P. On the demystification of mental imagery. *The Behavioral and Brain Sciences*, 1979, *2*, 535-581.

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.

Pinker, S., & Finke, R. A. Emergent two-dimensional patterns in images rotated in depth. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 244-264.

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

Shepard, R. N. Psychophysical complementarity. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

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.