ON THE ONTOLOGICAL STATUS OF VISUAL MENTAL IMAGES

Stephen Michael Kosslyn
Harvard University

There has long been considerable controversy over the ontological status of mental images. Most recently, members of the A.I. community have argued for the sufficiency of "propositional representation" and have resisted the notion that other sorts of representations are functional in the human mind. The purpose of this paper is to review what I take to be the best evidence that images are distinct functional representations in human memory. Before reviewing these data, however, I offer a preliminary definition of what I mean by a "visual mental image." This definition arises out of the "cathode ray tube" metaphor originally introduced in Kosslyn (1974, 1975, 1976) and later implemented in a computer simulation by Kosslyn & Shwartz (1977a, in press). On this view, images are spatial representations in active memory generated from more abstract representations in long-term memory; these spatial representations are able to be interpreted ("inspected") by procedures that classify them into various semantic categories.

1.0 A preliminary definition of a visual mental image

I wish to define a "visual mental image" in terms of five basic kinds of properties. Images are often distinguished from more discrete, propositional or linguistic representations because they supposedly have "analogue" properties. Thus, the first two properties noted below describe analogue representations as a class. Goodman (1968), Palmer (in press), Shepard (1975), Sloman (1975), and others have provided informative and detailed discussions of relevance here, and I will draw freely on these sources in the present discussion.

1) Images can capture continuous variations in shape. This continuity property implies that image representations are both semantically and syntactically "dense" or "undifferentiated" in the extreme (Goodman, 1968, p. 136 ff.). For example, a reading on a tire pressure gauge is an analogue representation to some extent, because every reading along the continuous scale has meaning (and so it is semantically dense); if the gauge had an infinity of markings of pounds-per-square inch, the scale would be syntactically dense and readings on it would be purely analogue. In contrast, discrete representations are not semantically or syntactically dense, but are differentiated (i.e., separable and distinct). For example, each reading of a digital clock, in contrast to the traditional dial variety, is entirely unambiguous in terms of its identity (i.e., is syntactically distinct) and its meaning (i.e., is semantically distinct). Images are both semantically and syntacically dense.

2) Part and parcel of the continuity property is the property that analogue representations are not arbitrarily related to their referents. Because analogue representations can be arranged on a continuum (e.g., of size), a symbol indicating a value falling between two others (e.g., an intermediate size) must refer to a value of the referent falling between the two indicated by the others (e.g., an object of intermediate size). Hence, unlike discrete representations, any given analogue representation cannot be assigned an arbitrary meaning (this point was first brought to my attention by Wilkins, 1977).

Because of this requirement, portions of images of surfaces or objects (involving two or three dimensions) bear a one-to-one structural isomorphism to the corresponding portions of the referent. That is, portions of the representation correspond to portions of the referent, and the spatial relations between portions of the referent are preserved in the image. This property has been described by Shepard (1975) as an "abstract first-order isomorphism." In

The work reported here was supported by NSF Grants BNS 76-16987 and BNS 77-21782. I wish to thank Willa Rouder for her assistance in preparing the manuscript.

this case, there is not a genuine first-order isomorphism, where a triangle is actually represented by something triangular in the brain, but there is a more abstract isomorphism where a triangle is represented by a set of representations corresponding to the vertices and sides standing in the proper relations. Thus, images depict, not describe. While any symbol can be used to represent an object or part thereof in a description, the particular representation of such in an image is constrained by other representations--given that the interportion spatial relations must be retained in the image representation.

The following three additional properties follow from our CRT metaphor:

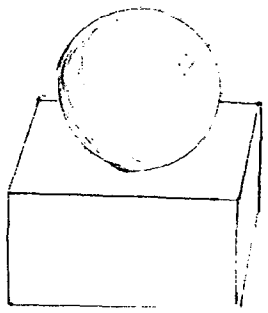
3) Images occur in a spatial medium that is equivalent to a Euclidean coordinate space. This does not mean that there is literally a screen in the head.[1] Rather, locations are accessed such that the spatial properties of physical space are preserved. A perfect example of this is a simple two-dimensional array stored in a computer's memory: There is no physical matrix in the memory banks, but because of the way in which cells are retrieved, one can sensibly speak of the inter-cell relations in terms of adjacency, distance, and other geometric properties.

4) The same sorts of representations that underlie surface images also underlie the corresponding percepts. Hence, in addition to registering spatial properties like those of pictures, images depict surface properties of objects, like texture and color. Thus, although the image itself is not mottled, or green, or large or small, it can represent such properties in the same way they are represented in our percepts. That is, the image representations must be able to attain states that produce the Qualia, the experience of seeing texture, color, size and so on.

5) Finally, by dint of the structural identity of image representations and those underlying the corresponding percept, images may be appropriately processed by mechanisms usually recruited only during like-modality perception. For example, one may evaluate an image in terms of its "size" (i.e., being depicted--the representation itself is neither large nor small) in the same way one would evaluate the representation evoked while actually seeing the object.

Images, then, share virtually all the properties of percepts, as opposed to properties of pictures or objects themselves. I refrain from making a complete identity because of a crucial difference: Perceptual representations are "driven" from the periphery, whereas images are somehow formed from memory. Hence, in both cases there may be particular kinds of "capacity limitations" that influence properties of the representation. For example (and this is an empirical question), images may be coarser and less detailed than the corresponding percept because of memory capacity limits.

These properties of images can be further understood in contrast to properties of "propositional" representations. Consider the two representations of a ball on a box illustrated in Figure 1. A propositional representation must have: 1) a function or relation; 2) at least one argument; 3) rules of formation; and 4) a truth value.



ON (BOX, BALL)

Figure 1. Two representations of a ball on a box.

In contrast:

1) Images do not contain identifiably distinct relations; relations only emerge from the conglomerate of the components being represented together. Thus, one needs two components before a relation like "on" can be represented.

2) Images do not contain arguments. The components of an image are not discrete entities that can be related together in precise ways. The box, for example, can be decomposed into faces, edges, and so on--and these are certainly not arguments in and of themselves.

3) Images do not seem to have a syntax (except perhaps in the roughest sense). That is, a relation like "on" requires two arguments in order to create a well-formed proposition; "on box" is an unacceptable fragment. In contrast, any syntax dictating "well-formedness" of pictures or images will probably depend on some sort of interaction with a "semantic component," will depend on what an image is supposed to be an image of. As we all know, "impossible pictures" are created regularly (e.g., by artists such as Escher), and rules that govern the nature of objects in the world may not

1. Although there could be, if images occur as topographic projections on the surface of the cortex; this kind of space is a subset of the one I am defining here, however.

necessarily constrain the things that one can depict in a picture.

4) Finally, unlike a proposition, an image does not have a truth value. In fact, as Wittgenstein (1953) pointed out, there is nothing intrinsic in a picture of a man walking up a hill that prevents one from interpreting it as a picture of a man sliding downhill backwards. The meaning of an image, and hence its truth value, are assigned by processes that work over the representation and are not inherent in the representation itself.

2.0 Five classes of empirical findings supporting the functional reality of visual mental images

2.1 Experiments on scanning visual images

A key property of images is that they embody spatial extent. If images are functional, then, we should expect this property to affect some forms of processing that involve using images. Kosslyn, Ball, & Reiser (1978) report a number of experiments that demonstrate that more time is required to scan further distances across mental images. In one study, people imaged a map containing seven locations and scanned between all possible pairs of locations. Time to scan increased linearly with increasing distance between the 21 possible pairs of locations, each of which was separated by a unique distance. There were no effects of distance in a control condition where subjects focused on a location in the image but then simply decided whether another object was present, without being asked to scan to that location.

In another experiment, people imaged schematic faces wherein the eyes were either light or dark and located either 3, 4, or 5 inches above the mouth; in all other respects the faces were identical. After a given face had been removed, a subject was asked to focus on the mouth and then to: image the face as large as possible without it seeming to overflow, or image it half of this size, or image it so large subjectively that only the mouth was left visible in the image. Following this, the word "light" or "dark" was presented. As soon as either word had occurred, the subject was to "glance up" to the eyes of the imaged face and see whether or not they were appropriately described by the word. Time to judge whether the eyes were light or dark increased linearly with distance from the mouth. Further, overall scanning times were reduced when people were asked to "shrink" an imaged face mentally prior to scanning it, and times were increased when subjects "expanded" a face before scanning. These results are difficult to explain if images are simple "abstract propositional" list structures, but follow naturally if images are spatial representations that preserve metric distance information.

2.2 Measuring the visual angle of the mind's eye

The notion that images embody spatial extent suggests that they may have spatial boundaries; after all, they do not extend on indefinitely. If images occur in a spatial representational medium, then their maximal spatial extent may be constrained by the extent of the medium itself. Kosslyn (in press) used the following paradigm in an attempt to test this idea: People were asked to image an object as if it were being seen from very far away. Then, they were asked to imagine walking towards the object and were asked if it appeared to loom larger; all subjects reported that it did (of the subjects who could do the task at all, which was usually only about 80% of the people tested). Further, these subjects claimed that the image loomed so large at one point that it seemed to "overflow." At this point, the subject was to "stop" in his/her mental walk and to estimate how far away the object would be if s/he were actually seeing it at that subjective size. We did this basic experiment in a variety of ways, having subjects image various sorts of pictures or image animals when given just their names and sizes; in addition, subjects estimated distance by verbally assessing feet and inches or responded by moving a tripod apparatus the appropriate distance from a blank wall.

If images occur in a spatially constrained medium, then the larger the imaged object, the further away it should seem at the point of overflow. In addition, a constant angle should be subtended by the imaged objects (which ranged in actual size) at the point of overflow. Using simple trigonometry, we were able to compute the "visual angle of the mind's eye" from the estimated distances and longest axis of each imaged object. In all of our experiments, the basic results were the same: First, people claimed that smaller objects seemed to overflow at nearer apparent distances than did larger objects (the correlation between object size and distance was always very high), and distance usually increased linearly with size of the imaged object. Second, the calculated "visual angle" at the point of overflow remained constant for different-sized objects when subjects imaged pictures or objects that had just been presented. The actual size of the angle varied, however, depending on instructions: More stringent definitions of "overflow" resulted in smaller angles. These last findings imply that images do not overflow at a distinct point, but seem to fade of gradually towards the periphery. (The best estimate of the maximal angle subtended by an image while still remaining entirely visible seemed to be around 20 degrees.)

In another experiment, we asked

people to scan images of lines subtending different amounts of visual arc and we calculated how many msec were required to scan each degree. These people also scanned an image of a line they had constructed to be as long as possible without either end overflowing. The visual arc subtended by this "longest possible non-overflowing line" was inferred from the time required to scan across it. This estimate was very close to one obtained using the technique described above and to one obtained by simply asking people to indicate the subjective size of a longest non-overflowing line by holding their hands apart so as to span the length of the longest line.

2.3 Effects of subjective size on ease of "seeing" parts of mental images

If asked which is higher off the ground, a horse's knees or the tip of its tail, many people claim to image the beast and to "inspect" the image, evaluating the queried relation. It makes sense to suspect, then, that images might be appropriately processed by the same sorts of classificatory procedures used in categorizing perceptual representations. If so, then we might expect constraints that affect ease of classifying parts perceptually also to affect ease of imagery classification. Parts of smaller objects are "harder to see" in perception, for example, and also may be harder to "see" in imagery. This result was in fact obtained (see Kosslyn, 1975); parts of subjectively smaller images of objects did require more time to classify mentally than did parts of subjectively larger imaged objects. In addition, simply varying the size of the part per se also affected time to examine an image. In this case, smaller parts--like a cat's claws--required more time to see on an image than did larger parts--like its head. This last result was obtained (Kosslyn, 1976) even though the smaller parts were more strongly associated with the animal in question, and were more quickly verified as belonging to the animal when imagery was not used (more highly associated properties are typically affirmed as appropriate more quickly than less associated ones in studies of "semantic memory"--see Smith, Shoben & Rips, 1974). These findings, then, not only are consistent with the notion that images are functional spatial representations that may be interpreted by other processes, but also serve to distinguish between processing imaginal and non-imaginal representations.

2.4 Effects of subjective size on later memory

If parts of subjectively smaller images are less distinct, then one might expect that the imaged object itself would be more difficult to identify. Thus, if one actually encodes a subjectively small image into memory, one's ability to recall the object later should be poorer than if the image had been larger--if in fact the image itself is recalled and inspected when one tries to recall the encoded words or objects. Kosslyn & Alper (1977) asked subjects to construct images of the objects named by pairs of words. Sometimes one of the images was to be very small subjectively and sometimes both images were to be "normal" sizes. When a surprise memory test for the words was later administered, memory was in fact worse if one member of a pair initially had been imaged at a subjectively small size. This result was replicated in several studies, each of which controlled for different possible confoundings (e.g., less "depth of processing" may have occurred when people constructed subjectively smaller images).

2.5 Transforming visual images

Cooper & Shepard (1973a, 1973b) and others have demonstrated that increasingly more time is required when one "rotates" a mental image through progressively greater arcs. Similarly, we have found that more time is required to expand or contract images to greater degrees (Kosslyn & Shwartz, 1977b), as did Sekular & Nash (1971). A propositional model of the sort offered by Gips (1974) does not lead us to expect these results. A spatial model, wherein a pictorial image is transformed, seems to imply in a straightforward manner that images will pass through intermediate positions as they are transformed, given that the same image is being retained and processed.

3.0 Conclusions

On my view, the most parsimonious, straightforward accounts of all these data will include the notion that images are functional representations in human memory. I have no doubt that alternative non-imagery accounts can be formulated for each set of results, but the collection of each of these individual accounts will likely be more ad hoc, post hoc and cumbersome than the imagery accounts.

References

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

Cooper, L. A. & Shepard, R. N. The time required to prepare for a rotated stimulus. Memory and Cognition, 1973b, 1, 246-250.

Gips. J. A syntax-derived program that performs a three-dimensional perceptual task. Pattern Recognition, 1974, Vol. 6, 189-199.

Goodman, N. Languages of Art: An Approach to a Theory of Symbols. Indianapolis, Indiana: Bobbs-Merrill, 1968.

Kosslyn, S. M. Constructing Visual Images. Ph.D. dissertation, Stanford University, 1974.

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

Kosslyn, S. M. Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval time. Memory and Cognition, 1976a, 4, 291-297.

Kosslyn, S. M. Measuring the visual angle of the mind's eye. Cognitive Psychology, in press.

Kosslyn, S. M. & Alper, S. N. On the pictorial properties of visual images: Effects of image size on memory for words. Canadian Journal of Psychology, 1977, 31, 32-40.

Kosslyn, S.M., Ball, T. M., & Reiser, B. J. Visual images preserve metric spatial information: Evidence from studies of image scanning. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 47-60.

Kosslyn, S. M. & Shwartz, S. P. Two ways of transforming mental visual images. Psychonomic Society Meetings, Washington D. C., 1977b.

Kosslyn, S. M. & Shwartz, S. P. Visual images as spatial representations in active memory. In E. M. Riseman & A. R. Hanson (Eds.), Computer Vision Systems. New York: Academic Press, in press a.

Palmer, S. E. Fundamental aspects of cognitive representation. In E. H. Rosch & B. B. Lloyd (Eds.), Cognition and Categorization. Hillsdale, N. J.: Lawrence Erlbaum Associates, in press.

Sekuler, R. & Nash, D. Speed of size scaling in human vision. Psychonomic Science, 1972, 27, 93-94.

Shepard, R. N. Form, formation, and transformation of internal representation. In R. Solso (Ed.), Information processing and cognition: The Loyola Symposium. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1975.

Sloman, A. Afterthoughts on analogical representation. Paper presented at the Conference on Theoretical Issues in Natural Language Processing. Cambridge, Mass.: June, 1975.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A feature model for semantic decision. Psychological Review, 1974, 81, 214-241.

Wittgenstein, L. Philosophical Investigations. New York: Macmillan & Co., 1953.