

Article

Thinking Visually

KRIS N. KIRBY AND STEPHEN M. KOSSLYN

Introduction

It often is convenient to divide visual information processing into two broad types. Low-level visual processes are driven solely by the stimulus input, and serve to detect edges, encode differences in wavelengths, establish depth using stereo disparity, and so on. In contrast, high-level visual processes involve the use of stored information. Such processes are utilized, for example, in the later stages of object recognition and identification and in navigation. In addition, high-level visual processes appear to underlie visual mental imagery (Farah 1988; Kosslyn 1987). In this chapter we explore some of the implications of the idea that visual mental imagery draws upon mechanisms used in visual perception. In particular, we consider visual thinking from the standpoint of high-level visual processing.

Speculation about the use of visual mental imagery in reasoning and problem solving has a long and rocky history (*e.g.* Boring 1950). The extreme views, represented by the British Associationists (who thought all thought relied on imagery) and the Behaviourists (who scoffed at the idea that such a thing could be studied) are well known. Even when prevailing philosophical and psychological attitudes have been generally sanguine about the susceptibility of mental events to scientific study, imagery has not always been in favour. Most recently, although information-processing approaches to cognition made the study of mental events generally respectable, the role of imagery as a distinct form of cognition has been regarded with skepticism in some quarters (*e.g.* Anderson and Bower 1973; Pylyshyn 1973, 1981).

The skepticism about imagery seems to have three roots. First, some have doubted the coherence of the very idea (*e.g.* Pylyshyn 1973; see Kosslyn and Pomerantz 1977, for a summary and review of these issues).

It seems fair to say that enough debate has now passed regarding these concerns to draw out the empirical issues and lay to rest the fundamental 'in principle' objections. Second, mental imagery, even if not an incoherent idea, seemed beyond the reach of science; for many years it was unclear how one could study such an ephemeral and ethereal creature in a rigorous way. However, new methodologies were developed, which allowed us to learn a lot about the role of visual imagery in memory (e.g. Paivio 1971, 1986), the similarities between images and like-modality percepts (e.g. Finke and Shepard 1986; Kosslyn 1987), and the properties of images as repositories of information (e.g. Kosslyn 1980; Shepard and Cooper 1982). Third, and perhaps most basic, there has been skepticism about why a linguistic animal would need imagery, and what it is used for. Indeed, there are some who claim never to have had the experience of mental imagery, and yet they appear to function adequately in the world.

One way of thinking about the purposes of imagery is to consider how and why it emerged during the course of evolution. From this perspective, it is useful to consider the relations between imagery and like-modality perception; this link may provide clues about the phylogenetic development of imagery. If, for example, visual imagery actually plays an important role in visual object recognition it is possible that imagery evolved primarily to serve this role. However, once a mechanism has developed for use in one domain, it can sometimes be taken over for use in other domains, a process called 'exaptation' (Gould and Lewontin, 1979). For example, it is thought that the earliest development of wings in insects was through a process of exaptation of appendages that had previously been used to cool the insect through a fanning action. Although these 'fans' evolved to help cool the insect, once they were in place new airborne niches became available, and mutations that allowed insects to exploit those niches were highly viable, which promoted the emergence of new species. Similarly, once visual imagery was available for use in high-level visual perception, it could then begin to be exapted for use in other domains, and ultimately in such sophisticated tasks as reasoning and problem solving. The use of visual imagery in visual thinking, then, would be a by-product of its role in visual perception.

For example, consider a recent series of experiments reported by Cave and Kosslyn (1989). Cave and Kosslyn were impressed by the computer vision system developed by Lowe (1987a, 1987b). Lowe noticed that once various bottom-up heuristics have been used to narrow down the possible identity of a stimulus to a few alternatives, it is useful to activate a stored 'model' of a likely object and project an image of it back to the input, comparing it with the stimulus. This process is particularly useful when the stimulus is partially occluded or otherwise obscured. To examine this idea, Cave and Kosslyn showed subjects rectangles and diamonds, and asked them to decide whether a form had sides of equal length. In some experiments two forms appeared overlapping, and the subjects were asked to evaluate the one drawn with light lines, whereas in other experiments only a single form was presented at a time.

One interesting result was that faster comparisons could be made when the same form appeared that had appeared on the previous trial, but this only occurred when it was embedded in a more complex stimulus configuration (containing another form or fragments of a form). Furthermore, Cave and Kosslyn varied the sizes of the stimuli, and found that more time was required when a stimulus appeared at a different size than the previous stimulus, with increasing amounts of time for greater disparities. This 'size scaling' function appeared both when the stimulus was the same as the one that appeared on the previous trial, and when the stimulus was different. However, when the target stimulus was embedded in a complex context—and only then—the slopes were very different in the two cases. When a form had to be recognized in a noisy environment, it appeared that subjects used an image of the previous form. Indeed, the size scaling rates found by Cave and Kosslyn were very similar to those reported by Larsen and Bundesen (1978) for scaling attention versus image size, suggesting that there were two mechanisms at work in the task—one that alters the scope of an 'attention window' (used when different figures appeared on successive trials) and one that alters the size of a mental image (used when the same figure appeared on successive trials). Cave and Kosslyn (1989) ruled out various alternative accounts for the results, and concluded that imagery does in fact play a special role in identifying objects in noisy environments, as predicted by Lowe's theory.

The notion that imagery emerged primarily in the service of perception leads to some interesting hypotheses about ways in which imagery might be most effective in visual thinking. When we ask about uses of visual imagery, we begin by asking what it is about imagery as part of the visual system that would make it suitable for use in various types of tasks. We then are led to ask about the kinds of tasks we would predict to benefit from its use. This paper is an attempt to provide a theoretical framework for studying the use of visual imagery in domains other than object recognition *per se*, particularly in the domains of reasoning and problem solving.

Thus, we begin by considering the properties of depictive representations, and how these properties make such representations useful in problem solving tasks. We then turn to key properties of the human visual imagery system, and will consider how these properties help determine when this system would be useful in problem solving. Finally, we discuss properties of classes of problems themselves, and how these properties of problems could help determine when imagery would profitably be used.

'Privileged Properties' of Depictive Representations

From a computational standpoint, perhaps the most salient characteristic of visual mental imagery is that it provides a mechanism for recalling

and manipulating depictive representations.¹ Before we can explore the possible utility of depictive representations, we must briefly characterize them. Depictive representations differ from propositional representations in both their syntax and semantics (Kosslyn 1984). Whereas the syntax of propositions allows for any number of symbols and classes of symbols to be combined according to a set of specific rules, depictive representations use 'points' as their sole class of symbols, and these are combined simply by spatial juxtaposition. And whereas the semantics of propositional representations depend solely on interpretation of its symbols (or sentences; Quine 1960), depictive representations convey meaning by visual resemblance. That is, each part of a depictive representation corresponds to part of the depicted object (or objects) such that relative 'distances' among the parts of the representation correspond to those among the actual parts themselves (see Kosslyn 1984).

The consequences of these properties of depictive representations will prove important for understanding the usefulness of visual imagery in a variety of domains. First, object shape properties are made explicit. All of the implicit local and global geometric relations among parts of the representation are accessible to varying degrees. For example, if asked to describe the shape of the empty enclosed space in an upper case letter A, one can generate an image and 'see' the triangle as easily as one can 'see' the shape of the lines. Or if one is asked 'What shape are a doberman's ears?', one might proceed to generate a visual image of the dog, 'inspect' its ears, and respond something like 'pointed'. To answer such questions using a propositional representation, one would have required the foresight to store this shape information explicitly in one's list of letter or doberman properties; the vast number of shape properties of parts of any normal object make it unlikely that one could have stored the explicit propositional information necessary to answer arbitrary questions such as these. Depictive representations make all of this spatial information available when needed.

A second, but closely related, consequence of using depictive representations is that spatial relations between objects are made explicit (this does not follow necessarily from the consideration just mentioned, *e.g.* if multiple objects could not be juxtaposed into the same array). This makes possible relative size judgments ('Was Hitler's moustache wider than Charlie Chaplin's?'), and relative position judgments ('Would President Bush be able to look an elephant in the eye while he "shook hands" with its trunk?'). These judgments can be either categorical, as in the two previous examples, or coordinate: requiring absolute metric information, as in

¹ An 'image', by definition, is depictive; the debate about imagery has hinged on whether the depictive properties of mental images that are evident to introspection are functional (see Kosslyn and Pomerantz 1977); the present chapter rests on the assumption that images are depictive and that these depictive properties can be functional, and explores the ways in which such representations could be exploited.

'Approximately how much longer is an American football than the diameter of a basketball?'

For present purposes there are two important implications of the above considerations: (1) depictive representations are likely to be particularly useful in any task requiring the assessment of shape or spatial information, and (2) depictive representations are an efficient way of storing spatial information. For any given object or scene, there is an enormous number of possible spatial relations that could be extracted, and it seems unlikely that a propositional representation large enough to store all of these relations explicitly could be managed, even if one had the foresight at the outset to encode all of the possible spatial relations. This is a clear example of a picture's being worth a thousand words, if not more.

Key Properties of the Human Visual Imagery System

Our observations about distinctive properties of depictions are suggestive, but the implications of these observations cannot be developed further unless we tie them to the structures and processes underlying human visual mental imagery. To gain insight into the types of tasks that would be facilitated by using imagery, we must consider the constraints and limitations of depictive representations used in human cognition. Central to our line of thinking is the idea that the visual imagery system provides a mechanism for recalling and manipulating depictive representations, and establishing mappings between non-spatial and spatial relations. We will focus now on these functions.

Structure: A Working Memory Store

Given the evidence that imagery and like-modality perception share brain mechanisms (*e.g.* Farah 1988), it is of interest that many visual areas in the brain are topographically mapped, and are capable of supporting depictive representations (Van Essen 1985). Indeed, Farah (1988) summarizes much data indicating that the occipital lobe is involved in imagery, which is known to contain numerous topographically mapped areas. Thus, when visual mental images are generated, we assume that the image representations proper are depictive and are maintained in a buffer; this buffer represents spatial information by preserving distances between points. We will refer to the array used in imagery as the 'visual buffer' to stress its connection with the visual system (Kosslyn 1987).

The idea that images occur in perceptual structures has far-reaching implications. For one, during perception it is important that visual areas have the neurological equivalent of a 'fast-fade phosphor' to avoid smearing of input during eye movements. The neural substrate does not 'automatically' retain a state for very long once the source of stimulation is removed. Thus, it is no surprise that information in the imagery array

decays rapidly over time. This property of imagery is of crucial importance when one considers possible uses of imagery. A primary constraint on the effective use of any memory store is its capacity limits, which determine the amount of information that can be held in the store at one time. The capacity limits of the visual buffer appear to have two sources.

First, because the neural substrate of the visual buffer does not automatically retain a state for very long after the input is removed, images decay very quickly. Hence, images can be maintained only through effort, by continually regenerating the input pattern (see Kosslyn 1980). There is a limit to how many entities one can retain at once in an image. For example, in an unpublished study (Kirby and Kosslyn 1990) eight subjects were asked to form a clear visual image of a brick, and to continue adding new bricks to the image until any of the bricks could no longer be seen clearly: on average, subjects reported being able to 'see' 6.3 bricks, with individuals ranging from four to nine. Consider an analogy to the number of balls one can juggle simultaneously: the capacity limit here is defined by the speed with which one can throw balls, the height one can throw them, and the speed with which they fall.

A second class of limitations on processing in the visual buffer is associated with the depictive nature of its representations: the represented objects can have only a limited scope (the size of the room in which one is juggling). Kosslyn (1978) and Finke and Kosslyn (1981) showed that objects can subtend only a limited visual angle before their edges begin to blur, as is expected if topographically mapped areas used in vision proper are also used in imagery (these areas typically magnify the fovea, allocating more than a proportional amount of cortex to input from it, and have decreased resolution toward the periphery). Depending on the type of object, the actual angle at which such blur is noticed varies. For objects with sharp edges, only about 20° of visual angle can be used effectively.

A third kind of limitation also concerns spatial factors: the visual buffer appears to have a 'grain'. When imaged objects subtend too small a visual angle, their details are difficult to 'see'. Indeed, Kosslyn (1976) pitted size and association strength of parts of objects against each other, and found that the time to 'see' a part of an imaged object was determined primarily by its size, whereas the time to recall a part not using imagery was determined by its association strength. (Kosslyn 1980 summarizes much data pertaining to the scope and resolution of imaged objects.)

Although it has not yet been investigated, it seems likely that, just as in visual perception, one must trade off detail for scope (Eriksen and St. James 1986; Shulman and Wilson 1987). By analogy to a video camera, as one zooms in to inspect finer details of the image, the portion of the image that remains in view must diminish. Conversely, for a given 'image distance' there is a limit to the detail that can be discerned on the imaged object. Unlike visual perception (at least beyond one's minimal focal length) there is an upper bound to the detail that is available in an image; one can only image to the detail that was stored in the long-term image

representation. For example, in perception one can inspect the detailed fur patterns on a doberman by moving physically closer and closer (up to some courage limit), but this may not be possible with a mental image: if one has never been close enough to a doberman to encode that visual information into memory, it cannot be included in a subsequent mental image.

Finally, if visual images do in fact occur in visual cortex, then a fourth limitation will exist: Images will compete for space with actual perceptual input. And in fact, there is much evidence that imagery and like-modality perception interfere with one another (e.g. Finke and Shepard 1986; Kosslyn 1980). This limitation often is easily overcome simply by closing one's eyes. But for some types of problems, one may need to compare an image to an actual object; in these cases, the limitations posed by competition for representational resources may be severe.

The visual buffer, then, has limited capacity, as determined by how many objects or patterns can (1) be retained at the same time, (2) fit within its limited scope, (3) subtend a large enough angle to be 'visible', and (4) be represented at the same time within the structure without interfering with each other. The limitations on amount, scope, and size are partially interdependent. To pursue our juggling analogy, one might have the ability to juggle a large number of objects but be limited to less than one's maximum because of the size of the room in which one is juggling. Similarly, if the objects are too big, they could not be tossed about in a relatively small room, but if they are too small they may be difficult to see. Thus, the two spatial constraints are clearly interdependent: imaged objects cannot subtend too large an angle or they will seem to overflow, but cannot subtend too small an angle or they will not be identifiable.

The limitations on the amount of material that can be retained at the same time appear to be best characterized in terms of shape units ('chunks'). It is clear that the time to form an image hinges on the number of perceptual units, not the actual physical properties of the stimulus. For example, if subjects are asked to see a Star of David as two overlapping triangles when studying it, they later can form a visual mental image of the star much faster than if they are asked to see it as a hexagon with six small triangles along its sides. Similarly, if subjects see an array as three rows of six letters, they later visualize it more quickly than if they see it as six columns of three letters (Kosslyn 1980). We assume that this generation process is critical in maintaining images, given that the buffer itself does not retain patterns of activation. Thus, it is of interest to note that although only a handful of perceptual units can be maintained in an image at once (Weber and Harnish 1974; Kosslyn, Brunn, Cave and Wallach 1984), this limit does not have a simple relation to the number of spatial relations that can be maintained. Because the number of spatial relations among a set of units is not a simple function of the number of units, there is no direct limit on the representation of spatial relations *per se*, only on the number and complexity of the units being related. For example, with only

two units (e.g. one unit two inches Northwest of the other) one could encode the relations left-of, right-of, above, below, collinear, two inches from, Northwest, and so on. The number of spatial relations increases exponentially with the addition of a third unit.

The level of performance of a task relying on imagery will decrease dramatically when a capacity limit is reached. This can occur when the information required to solve a problem exceeds the limit that one can maintain in an image at once, or when a concurrent memory load placed on the system by a secondary task effectively reduces the limit. For example, imagine trying to determine whether a race horse's knees are above the tip of its tail while concurrently imaging a route through a maze on a computer screen. To the extent that the latter information taxes the representational capacity of the visual buffer or diverts effort necessary to maintain the image of the horse, it will become increasingly difficult to use imagery to access the information relevant to the question.

The constraints imposed by the capacity limits of the visual buffer in some cases might be overcome by using a second, longer-term, memory store as a 'swap-space', moving information in and out of the visual buffer as it is needed. The use of imagery in visual thinking can be thought of as an instance of the use of 'working memory'. Information in working memory corresponds to information that has been activated and is available for immediate use in problem solving. Working memory typically is conceived as consisting of a central executive and at least two independent stores, the visual/spatial and the articulatory/phonological (Baddeley, 1986). Each of the stores presumably has its own capacity limit. It is possible that when one of these stores, perhaps the one normally used in a given task, reaches its capacity limit the other store could be used as an additional temporary store, or 'swap space' for the primary store. This conjecture apparently applies to memory *per se*; Kosslyn, Holyoak, and Huffman (1976) showed that people could retain more words if they imaged some and verbalized some, rather than using only a single method for the entire list. Similarly, if one tries to retain both a telephone number and a street address long enough to cross the room and write them down, it may be necessary to rehearse the phone number in the articulatory store and 'picture' the street address (the process of writing it down may interfere with the image, however, so write down the image information first!). This type of application does not appear to be widespread, and may depend on learning specific control strategies for coordinating the swapping. However, this process would take measurable time and increase the probability of error because the image generation process is not flawless.

Thus, the limited capacities of the visual buffer will affect the difficulty of tasks when either the amount of information needed to solve a problem exceeds capacity, or when part of this capacity is taken by a concurrent task. To the extent that the capacity limit can be circumvented by using another memory store as a swap space, the time to solve the problem will be substantially lengthened and the probability of error increased.

However, a 'chunking' mechanism that forms new hierarchical relations within a depiction should functionally expand the limits on the visual buffer, grouping portions of the image into more abstract structures that can be refreshed as wholes. For example, if one tries to generate an image of the night-time sky it would be very difficult to generate or refresh images of a significant number of individual stars even though they represent the simplest possible chunks: single points. However, if one were to group stars into constellations these constellations would themselves become chunks and could be refreshed as wholes. These new chunks would be more abstract, in that the individual star information would no longer be explicit, but finer-grained information could be accessed by 'zooming in' on a particular constellation. The chunking in this example is most likely to occur during visual perception, while one is actually looking at the sky, but a similar process could occur while one is 'looking' at an image of a number of disparate items (Finke, Pinker and Farah 1989). This process would allow the imagery system itself to increase the amount of information it can work with within the confines of the spatial and temporal limits on the visual buffer.

Processes: Manipulating Chunks

So far we have focused primarily on the usefulness of the visual imagery system in tasks requiring recall and inspection of visual and spatial information. Yet within the imagery system spatial relations and shapes can be manipulated and reinspected, and these functions greatly enhance the kinds of problems to which imagery can be applied. The various kinds of manipulations have received a great deal of attention in recent imagery research. Rather than simply recount well known findings (for a review, see Shepard and Cooper 1982), we will try rather to illustrate how some of the most well-documented kinds of image manipulations could play a role in some problem solving tasks.

Rotation. Suppose one were asked whether a lower case *p* is the same as an upside-down *b* or *d*. Many people report forming a mental image of a *p*, mentally rotating it 180°, and simply inspecting the image for the answer. This introspection receives support from the results of formal experimentation: rotation takes measurable time and has been much investigated by Roger Shepard and his colleagues. It is not difficult to describe more 'ecologically valid' situations in which such an ability might be used. For example, suppose that you have just walked through a complex building for the first time, memorizing the floor plan of the path you took as you went, and now wish to return the way you came. One simple strategy is to generate an image of the floor plan and mentally rotate it 180°, using the inverted version to guide your walk back. Another example, if not everyday task, in which rotation would be useful is working jigsaw puzzles. One starts with a pattern and looks for its complement in a large

array of scattered shapes. The task is eased if one can generate an image of the 'missing' shape and rotate it to compare it with the available shapes, rather than picking up each piece and trying it out manually. A conceptually similar algorithm was used by Lowe (1987a) in a computer program that could pick out the individual disposable razors in a bucket of jumbled, partially overlapping razors.

Translations. An object in an image can be translated from one position to another. This is an important property because it permits objects to be juxtaposed in different ways for comparison. For example, suppose that you were replacing the back on a broken wooden chair, where the back had pulled out of square holes in the seat. You measure the holes in which the original pieces had been inserted and determine that they are two inches across the diagonal. At the furniture store the clerk tells you that they no longer make those pieces, but now use dowels whose ends form two-inch diameter circles. Will these new pieces fit into the old seat? One strategy for answering this question is to generate an image of a two inch diagonal square and a two inch diameter circle and juxtapose them so that their compatibility can be determined. By doing this one could readily 'see' that the circle would intersect the square at all four corners and therefore would not fit inside.

As a second example, suppose that you have repaired your chair and now wish to rearrange the furniture in the room. You want to know whether the sofa will fit against the wall. Rather than measuring the sofa and the wall, or pushing the sofa into place, one could generate an image of the sofa against the wall and 'see' whether it fits. This will only work to the degree to which the veridical relative sizes of the sofa and wall can be preserved in the image, and there is considerable room for error when the sizes are close. However, this strategy would suffice in a great many situations and save one considerable stress on the lower back.

Size alteration. When one generates an image of an object one is not stuck with considering that object at the size evident in the image. Objects in an image can be expanded or contracted. For example, when you determined earlier that the two-inch diameter dowels would not fit into the two inch diagonal holes, you might wonder how much the dowels would need to be cut down to make them fit. By keeping your image of the square constant and shrinking the size of the circle you could get a rough estimate of how small the circle would have to be to fit. Or suppose that you were buying a T-shirt for a friend whose shirt size you did not know. You might hold up one T-shirt and imagine it expanding or shrinking until it seemed to fit the size of your friend in the image.

Some of us often buy shirts of the wrong size for our relatives and friends, and the point of these examples is not to argue that imagery is a precise method for making spatial comparisons. The value of imagery is that it offers a low-cost heuristic for making such comparisons, and because

of this feature, its use is pervasive in everyday situations.

Motion. Perhaps the most intriguing, and least studied, property of the imagery system is its ability to use sequences of the above three transformations to perform analog simulations of real motion. In these cases, images represent metric spatial relations over time. This gives the organism the ability to preview courses of action, or forecast outcomes of events currently underway in its environment. We can distinguish among different cases of this use of imagery, emphasizing the primary aspect of the tasks, but it should be noted that each borrows from the others to some degree. They are perhaps best illustrated by concrete examples.

1. Extrapolation of current movement. (a) Will I need to slow my car to allow another car on the entrance ramp to enter my lane? (b) In football, for a quarterback to properly 'lead' a receiver when throwing a pass he needs to judge both the speed and course of the receiver (extrapolating his current movement) and see to it that the ball crosses that course at an appropriate time.
2. Extrapolation of size alteration. Can I carry this box through that door? One may (a) imagine oneself as though watching another person from behind, moving forward and trying to pass through the doorway with the box, or (b) simply imagine the doorway approaching, taking into account how the doorway would 'grow' as it approached.
3. Process simulation. (a) What would happen if I were to push this rolling table down the hall into that stack of folding chairs on top of which the cat is sleeping who when startled always jumps out of the window with the flower pot on the sill? (b) When assembling or installing a new appliance, say an air conditioner, most of us have had the experience of finding that the line drawings in the instructions usually do not map precisely onto the pieces that actually come in the box. Before even holding the pieces up for comparison one might run a quick imagery simulation of a few steps to see if there is any chance of the pieces fitting together as desired. (c) People probably cannot solve the entire Tower of Hanoi problem in their heads, but they can work ahead a few steps, testing actual sequences of moves by simulating them using visual imagery. One might, in fact, be able to solve the 'Missionaries and Cannibals' problem totally by imagery simulation if one has enough patience. Finally, there is some evidence that chess players use visual imagery to 'think ahead' a few moves during play (Holding 1985).

The ability to perform mental simulations of ongoing activities has important implications for skill acquisition. When the goal of skill learning is to string together a set of disparate movements, it is important to rehearse those particular movements in the correct order without intrusion

from unwanted movements. If a mental simulation of the activity is being run slightly in advance of the actual behavior, one could test each move before actually making it, and thereby limit the actual behavior to only the desired, correct movements. This makes virtually error-free performance possible even on relatively unlearned tasks, and might significantly enhance the learning of a particular pattern of movement. And in fact, there is much evidence that imagery can enhance such learning (e.g. Richardson 1969).

Non-spatial to Spatial Mapping

The foregoing examples have primarily focused on the use of images of objects or scenes as they would be seen in visual perception. But the imagery system is much more versatile than this: correspondences between non-spatial information and spatial representations can often be established, and imagery can be used to solve non-spatial problems. Thus, symbols (letters, numbers, tokens, etc.) can be manipulated in the visual buffer in the same way that they are manipulated on paper. For example, try doing a simple three digit addition problem (e.g. $479 + 187$) in one's head by simulating the manner you would do it on paper. Although the addition of each pair of digits is a reflexive process for most of us, for the two rightmost columns one must 'carry the ones' to the next column to the left. For most people, these new digits can be easily encoded into an image of the numbers above the appropriate columns. (By this point, some readers will have now discovered that individual differences in imagery are more relevant to personal concerns than had heretofore been realized.)

The most extensively studied application of this type of symbol manipulation is in the study of syllogistic reasoning. When solving transitive inference problems, or 'three-term syllogisms', such as 'Joe is smarter than Bill; Bill is smarter than Tom; who's smartest?', many people report that they image symbols representing each person in the problem (perhaps the first letter of each name, or simply a dot, for example) and arrange them in a linear visual array corresponding to their order on the 'intelligence' dimension. Empirical data support these reports (Huttenlocher 1968; Shaver, Pierson and Lang 1974; Ormrod 1979). In particular, Huttenlocher (1968; Huttenlocher *et al.* 1970) showed that the relative difficulty of many of these problems corresponds to the relative difficulty of placing real objects in a linear array.

In a related paradigm, Johnson-Laird (1983) suggests that people solve classical syllogisms by concretizing the items in the premises and then manipulating these representations. For example, a person might represent 'Some dogs are black' by generating representations ('tokens') that stand for particular black dogs, and then might manipulate the relations between these tokens and the tokens from another premise in order to determine whether a given conclusion is warranted. Johnson-Laird's experiments show that the order of problem difficulty can be predicted by the number

of different relationship structures ('models') that must be made to test whether a given conclusion is consistent with a given pair of premises.

Although Johnson-Laird does not take this step, it is straightforward to suggest that many people perform this token manipulation by generating images of the tokens in the visual buffer, and represent relations among tokens by spatial position in the array. In fact, in his illustrations of this process, Johnson-Laird (1983) represents the relationships among tokens by spatial position on the page. Given the capacity limits of the system, one would expect that syllogisms would become more difficult when the number or complexity of the models required to solve them approached or exceeded these limits. This suggests that one's ability to solve syllogisms in one's head might be predicted by measures of imagery capacity.

Problem-Solving Tasks that Engender Imagery

We are now in a position to outline classes of problems in which visual imagery is likely to play a functional role. Such an analysis is useful because it can help us acquire a better understanding of the tasks themselves, and because of the empirical implications for the use of imagery in the tasks. In general, imagery should be useful when (a) a problem addresses (or can be made to address) visual properties that do not exceed the visual buffer's capacity limits, or (b) spatial relations are critical (or can be made critical) for deriving the solution. Depictive representations can be used in an abstract problem if a mapping can be achieved between the elements in the problem and objects in an image. If an effective mapping can be established between the problem and a depictive representation, visual mental imagery should play a useful role.

Thus, the generality of imagery as a means of solving problems hinges on the ability to map problems into depictions. What follows are two classes of problems that are defined according to the principles that determine how they can be depicted. 'Direct' mapping is the depiction of visual properties of an actual object or objects in the image (although perhaps in an abbreviated or idealized form). In contrast, 'abstract' mappings require a second layer of interpretation: the depiction stands for something else (e.g. a dot for a person, a statue holding scales for justice, spatial proximity for class inclusion). In cases of the latter type the complexity of the mapping rule adds another set of factors affecting the ease of using imagery. Thus, the challenge in these problems is to devise a depiction that is effective not only because it respects the capacity limits of the imagery system but because it is unambiguous and easy to interpret symbolically.

Direct Depiction

The class of problems most likely to command the use of the visual imagery system are those that require one to consider subtle spatial relations or

visual properties of objects not present. Examples of this type were given above regarding dobermans' ears, Hitler's moustache, and footballs. In each of these cases it is not likely that the required visual or spatial information was previously stored explicitly in one's long-term memory. Rather, the information was stored implicitly, and extracted from the image in the same way that it would be from the actual stimulus.

Problems of geometry are well suited for imagery, and fit in this class. For example, if asked what shape lies at the center of a Star of David, most people would need to form an image of the star and inspect the center. Geometry problems in two dimensions are subject to the capacity limits discussed above, and it is probably the case that the limit on three-dimensional problems is even more strict. Questions that require the use of subtle relations among purely visual properties are also likely to invoke imagery. For example, most people report using imagery when deciding whether a Christmas tree is a darker green than a frozen pea (Kosslyn 1980; Kosslyn and Jolicoeur 1981). (Note that whenever we consider a situation to be a 'problem', we are assuming it has not been dealt with in the past by the person; if it has, the answer typically can be reproduced via rote memory, retrieving information stored propositionally.)

Abstract Mapping

A second class of problems does not require one to consider spatial or visual information *per se*, but recruits imagery by establishing a mapping between non-spatial information and spatial representations. There are two situations in which such a mapping would be useful. The first is when a depictive representation would be more concise than a verbal or propositional representation alone. For example, finding a representation that facilitates solution of the problem is the impetus behind setting up and working with 'models' of syllogisms rather than carrying out a purely syntactical calculation. Just as Venn diagrams were developed to facilitate the solution of syllogisms on paper, people may spontaneously adopt depictive representational methods for reasoning problems performed 'in the head'. In these cases, the depictions should have a direct correspondence between spatial and conceptual properties. For example, putting one circle inside another corresponds directly to the idea that members of one class are also members of another class; similarly, having more of a visual dimension correspond to more of a quantity (*e.g.* increased height for increased intelligence) has such a direct correspondence. Thus, the depiction is useful not only because it is concise, having fewer 'chunks' than would be required in a different format, but because there is a simple and direct way of decoding the depiction.

The second situation is when the imagery method of solution involves a simpler algorithm. For example, consider a problem in which one tries to 'exclusive-or' (XOR) together two strings of binary digits. The exclusive-or takes a pair of digits (*e.g.* 1, 0) and yields a 1 if either the first *or* the

second digit (but not both) is a 1, and yields a zero otherwise. To XOR together two strings, one must consider each pair of digits corresponding in serial position in the two strings, XOR them, and keep track of a third string containing the results of XOR'ing these pairs. Now try XOR'ing together the strings '10110' and '01110' without looking back at the strings while you work. This is a very difficult task to perform in one's head, but most of our pilot subjects found it tractable when they generated a visual image of the two strings, one above the other, and visualized the solution below. As the strings increase in length, a more efficient algorithm is to image the two strings, then inspect each pair of digits serially while subvocally rehearsing (acoustically imaging) the result string as it is produced. It is difficult to devise a plausible algorithm for this task that can be performed without external memory aids and avoids using a depictive representation.

Conclusions

If we assume that image representations are depictive, we are led to infer that three properties of imagery can be exploited to aid problem solving. First, visual information can be effectively organized into 'chunks'. Thus, one can reason about a large configuration of stars if they are treated as a constellation, or can reason about complex geography if an imaged map is organized into neighborhoods. Such chunking is necessary to circumvent the notorious capacity limitations of imagery. Second, spatial relations are in a sense 'free' in depictions. Unlike propositional representations, where such relations must be explicitly noted, in depictions spatial relations are an emergent property of sets of depicted perceptual units. Thus, to the extent that such relations can be used to represent key information, or manipulations of such relations can be used to solve problems, depictions will be useful. Finally, depictions need not be 'literal'. Depicted patterns can stand for other things. To the extent that the first two properties can be exploited, the effectiveness of imagery in these cases will be determined by the ease of using the decoding rules. If analogies can be exploited (*e.g.* 'inside' conceptually depicted by a shape inside another), such rules can be easily learned and used.

What is missing here are precise and general rules for how to decide when it is a good idea to use imagery to solve any given problem. One major reason such rules are missing is that they probably are impossible to formulate, for two reasons. First, the range of individual differences in imagery abilities is very large. To the extent that effective imagery hinges on working within capacity limitations, we cannot prescribe imagery for any given problem: people differ widely in their imagery capacities (Kosslyn *et al.* 1985). Second, the same problem applies to decoding rules when non-spatial mappings are used: what is easy for one person will not be for another. One reason a rule is easy to use is that one's previous

knowledge allows one to learn the rule quickly and to remember it easily. People differ widely in their knowledge, and hence we cannot expect a single dictum to apply universally.

Nevertheless, imagery has useful properties that might be exploited in ways currently overlooked. Our goal in this chapter was to explicate the relevant variables, and point out ways in which properties of imagery lend themselves to certain types of problem solving. We suspect that if it is possible to characterize the strengths and weaknesses of a given individual's imagery processing system and to characterize his or her range of knowledge in an area, it may be possible to develop more precise rules for the use of effective imagery in that area. By realizing that imagery is part of the visual system we can show how some properties of the visual system make predictions about properties of imagery, and it will be interesting to see in the future how the study of imagery benefits from further discoveries about the visual system as a whole. It might seem only fitting if a full understanding of imagery required an understanding of broader areas of psychology; after all, imagery has not only often been viewed as at the fringe (or outside) of psychology, but also has often been viewed as at the very heart of psychology. In our view it probably is somewhere in between, and a full understanding of imagery will both draw from and inform other aspects of psychology.

*Department of Psychology
Harvard University
33 Kirkland Street
Cambridge, MA 02138
USA*

Author Notes

This work was supported by a grant from the AFOSR (88-0012) to the second author. Requests for reprints should be sent to the first author at the Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA, 02138.

References

- Anderson, J. R. and Bower, G. H. 1973: *Human Associative Memory*. New York: V. H. Winston and Sons.
- Baddeley, A. 1986: *Working Memory*. New York: Oxford University Press.
- Boring, E. G. 1950: *A History of Experimental Psychology*, 2nd edition. New York: Appeltan-Century-Crofts.

- Cave, K. R. and Kosslyn, S. M. 1989: Varieties of Size-specific Visual Selection. *Journal of Experimental Psychology: General*, 118, 148–64.
- Eriksen, C. W. and St. James, J. D. 1986: Visual Attention Within and Around the Field of Focal Attention: A Zoom Lens Model. *Perception and Psychophysics*, 40, 225–40.
- Farah, M. J. 1988: Is Visual Imagery Really Visual? Overlooked Evidence from Neuropsychology. *Psychological Review*, 95, 307–317.
- Finke, R. A. and Kosslyn, S. M. 1980: Mental Imagery Acuity in the Peripheral Visual Field. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 126–39.
- Finke, R. A., Pinker, S. and Farah, M. 1989: Reinterpreting Visual Patterns in Mental Imagery. *Cognitive Science*, 13, 51–78.
- Finke, R. A. and Shepard, R. N. 1986: Visual Functions of Mental Imagery. In K. R. Boff, L. Kaufman and J. P. Thomas (eds.), *Handbook of Perception and Human Performance*. New York: Wiley-Interscience.
- Gould, S. J. and Lewontin R. 1979: The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society, Series B*, 205, 581–98.
- Holding, D. H. 1985: *The Psychology of Chess Skill*. Hillsdale, NJ.: Erlbaum Associates.
- Huttenlocher, J. 1968: Constructing Spatial Images: A Strategy in Reasoning. *Psychological Review*, 75, 550–60.
- Huttenlocher, J., Higgins, E. T., Milligan, L. and Kaufman, B. 1970: The Mystery of the 'Negative Equative' Construction. *Journal of Verbal Learning and Verbal Behavior*, 9, 334–41.
- Johnson-Laird, P. N. 1983: *Mental Models*. Cambridge, MA.: Harvard University Press.
- Kirby, K. N. and Kosslyn, S. M. (1990). [Capacity Limits on Visual Images of 'Bricks'.] Unpublished data.
- Kosslyn, S. M. 1976: Can Imagery be Distinguished From Other Forms of Internal Representation? Evidence From Studies of Information Retrieval Time. *Memory and Cognition*, 4, 291–7.
- Kosslyn, S. M. 1978: Measuring the Visual Angle of the Mind's Eye. *Cognitive Psychology*, 10, 356–89.
- Kosslyn, S. M. 1980: *Image and Mind*. Cambridge, MA.: Harvard University Press.
- Kosslyn, S. M. 1984: Mental Representations. In J. R. Anderson and S. M. Kosslyn (eds.), *Tutorials in Learning and Memory: Essays in Honor of Gordon Bower*. New York: W. H. Freeman.
- Kosslyn, S. M. 1987: Seeing and Imagining in the Cerebral Hemispheres: A Computational Approach. *Psychological Review*, 94, 148–75.
- Kosslyn, S. M., Brunn, J. L., Cave, K. R. and Wallach, R. W. 1984: Individual Differences in Mental Imagery Ability: A Computational Analysis. *Cognition*, 18, 195–243.
- Kosslyn, S. M., Holyoak, K. J. and Huffman, C. S. 1976: A Processing Approach to the Dual Coding Hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 223–33.
- Kosslyn, S. M. and Jolicoeur, P. 1981: A Theory-based Approach to the Study of Individual Differences in Mental Imagery. In R. E. Snow, P. A. Federico, and W. E. Montague (eds.), *Aptitude Learning and Instruction: Cognitive Processes Analysis of Aptitude, Vol. I*. Hillsdale, NJ.: Erlbaum Associates.

- Kosslyn, S. M. and Pomerantz, J. R. 1977: Imagery, Propositions and the Form of Internal Representations. *Cognitive Psychology*, 9, 52-76.
- Larsen, A. and Bundesen, C. 1978: Size Scaling in Visual Pattern Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 1-20.
- Lowe, D. G. 1987a: Three-dimensional Object Recognition from Single Two-dimensional Images. *Artificial Intelligence*, 31, 355-95.
- Lowe, D. G. 1987b: The Viewpoint Consistency Constraint. *International Journal of Computer Vision*, 1, 57-72.
- Ormrod, J. E. 1979: Cognitive Processes in the Solution of Three-term Series Problems. *American Journal of Psychology*, 92, 235-55.
- Paivio, A. 1971: *Imagery and Verbal Processes*. New York: Holt, Rinehart and Winston.
- Paivio, A. 1986: *Mental Representations*. New York: Oxford University Press.
- Pylyshyn, Z. W. 1973: What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psychological Bulletin*, 80, 1-24.
- Pylyshyn, Z. W. 1981: The Imagery Debate: Analogue Media versus Tacit Knowledge. *Psychological Review*, 87, 16-45.
- Quine, W. V. 1960: *Word and Object*. New York: Wiley.
- Richardson, A. 1969: *Mental Imagery*. New York: Springer Publishing Company.
- Shaver, P., Pierson, L. and Lang, S. 1975: Converging Evidence for the Functional Significance of Imagery in Problem Solving. *Cognition*, 3, 359-75.
- Shepard, R. N. and Cooper, L. A. 1982: *Mental Images and their Transformations*. Cambridge, MA.: MIT Press.
- Shulman, G. L. and Wilson, J. 1987: Spatial Frequency and Selective Attention to Local and Global Information. *Perception*, 16, 89-101.
- Van Essen, D. 1985: Functional Organization of Primate Visual Cortex. In A. Peters and E. G. Jones (eds.) *Cerebral Cortex (Vol. 3)*. New York: Plenum Press.
- Weber, R. J. and Harnish, R. 1974: Visual Imagery for Words: The Hebb Test. *Journal of Experimental Psychology*, 102, 409-14.