THE BEHAVIORAL AND BRAIN SCIENCES (1979) 2, 535–581
Printed in the United States of America

On the demystification of mental imagery

Stephen M. Kosslyn
Department of Psychology and Social Relations, Harvard University, Cambridge, Mass. 02138

Steven Pinker
Department of Psychology and Social Relations, Harvard University, Cambridge, Mass. 02138

George E. Smith
Department of Philosophy, Tufts University, Medford, Mass. 02155

Steven P. Shwartz
Department of Psychology, The Johns Hopkins University, Baltimore, Md. 21218

Abstract: What might a theory of mental imagery look like, and how might one begin formulating such a theory? These are the central questions addressed in the present paper. The first section outlines the general research direction taken here and provides an overview of the empirical foundations of our theory of image representation and processing. Four issues are considered in succession, and the relevant results of experiments are presented and discussed. The second section begins with a discussion of the proper form for a cognitive theory, and the distinction between a theory and a model is developed. Following this, the present theory and computer simulation model are introduced. This theory specifies the nature of the internal representations (data structures) and the processes that operate on them when one generates, inspects, or transforms mental images. In the third, concluding, section we consider three very different kinds of objections to the present research program, one hinging on the possibility of experimental artifacts in the data, and the others turning on metatheoretical commitments about the form of a cognitive theory. Finally, we discuss how one ought best to evaluate theories and models of the sort developed here.

Keywords: computer simulation; imagery; memory; mental representation; perception; visual information processing

A history of mental imagery would almost require a complete history of the idea of mental representation, so intimate is the relationship between the two concepts. The objections to mental imagery have traditionally been of two forms. First, it has been argued that imagery cannot serve the functions that have been attributed to it. Most notably, it has been pointed out (at least since Berkeley's time) that an image cannot represent an object or scene uniquely without some interpretive function that picks out certain characteristics of the image as being important and others as being incidental. That is, an image of John sitting could represent John, John's head, bent knees, and so forth, depending on what one pays attention to in the image. And the "stage directions" indicating what is important in an image cannot *themselves* be images – if they were, the problem would only be pushed back a step. This class of objections is to the point: images cannot be the *sole* form of internal representation that exists in human memory. But this does not mean that images cannot be *one* form of representation in memory.

The second class of objections historically leveled against the use of mental imagery as an explanatory construct in psychology has two thrusts: first, it has been claimed that there are incoherencies and inconsistencies inherent in the concept. Pylyshyn (1973) has recently summarized and developed these claims, and Kosslyn and Pomerantz (1977) have provided counterarguments. Not surprisingly, neither the arguments nor the counterarguments have been definitive, and neither seems to have had enough force to sway most people from whatever position they found most congenial in the first place. In the present paper we will not attempt to argue from purely rational grounds that mental imagery is a suitable topic for psychological study and a suitable explanatory construct in psychology. Rather, our argument will consist of a demonstration that progress can in fact be made in studying imagery scientifically. The second thrust of these objections against the use of imagery as an explanatory construct focuses on the claim that imagery is not a well-formed domain in its own right, but is merely one special aspect of a more general processing system (see Pylyshyn 1973). Again, if this were the case, one would not expect to see much progress in attempts to develop a special theory of imagery. However, if a coherent theory that treats imagery as a distinct "mental organ," a theory having explanatory power and predictive utility, *can* be developed, this alone should make us hesitate to abandon the construct. In the course of describing the theory and its development we will raise questions about how imagery – or any other mental structure or process – ought to be studied and how theories of mental phenomena ought to be evaluated.

This paper has three main sections. In the first we outline some particulars of the approach to theory construction that is adopted here and that has guided the research program since its inception. In addition, we present an overview of the empirical foundations of the theory, briefly describing four issues that we attempted to resolve empirically before beginning to construct a detailed model. In the second section, we present the core theory itself and describe how it has been instantiated in a computer simulation model. We will discuss not only the model itself, but the rationale for using a computer simulation model per se. Finally, we conclude by considering a number of possible issues, problems, and objections surrounding the present program.

 0140-525X/79/KOPSS 029/$04.00

1.0 Empirical foundations of the present theory

The present research program had two phases. In the first, we attempted to delimit empirically the class of acceptable models. We began with a simple conception of how the imagery representation system might operate. This conception hinged on the notion that visual images might be like displays produced on a cathode ray tube (CRT) by a computer program operating on stored data. That is, we hypothesized that images are temporary spatial displays in active memory that are generated from more abstract representations in long-term memory. Interpretive mechanisms (a "mind's eye") work over ("look at") these internal displays and classify them in terms of semantic categories (as would be involved in realizing that a particular spatial configuration corresponds to a dog's ear, for example). This simple "protomodel" was used as a heuristic to help construct a "decision tree" in which the nodes represented issues and the branches stood for alternative positions on the issues. Sets of experiments were conducted to eliminate branches (as far as possible), allowing us to descend to the next issue. The decision tree we ultimately formed is illustrated in Figure 1.

Our CRT protomodel directed our attention to the following four key issues: first, it suggested that the "quasi-pictorial" image we experience is not an epiphenomenal concomitant of more abstract, nonpictorial processing; second, it led us to ask whether such images are simply retrieved or can be generated; third, if images are generated, we could then ask whether generation is simply a piecemeal retrieval of stored information, or whether it involves retrieving organized units; last, we were faced with the question of whether images are composed solely by retrieving encodings of how something appeared (the products of "seeing as"), or whether "descriptive" information (such as the products of "seeing that") is also used. At the end of Phase I, then, we had a set of constraints on the viable data structures and processes of a theory of imagery. Let us now briefly review the progress of empirical work in Phase I (see Kosslyn, in press, for more detail).

1.1 Issue I: are images epiphenomenal?

The CRT metaphor posits that "quasi-pictorial" images are produced and then processed by other mechanisms. Such an image is not strictly pictorial because it does not share all the properties of pictures (e.g., it cannot be hung on a wall). Rather, it is quasi-pictorial in that it *depicts* information, as opposed to *describing* information in a discursive way. Presumably, information implicit in long-term memory becomes explicit in an image (e.g., people claim that when asked which is higher off the ground, a horse's knees or the tip of its tail, the information becomes apparent only when they form an image of the beast). Alternatively, images could be nonfunctional, epiphenomenal concomitants of more abstract unconscious processing. On this view, images could simply be like the lights that flash on the outside of a computer while it is adding; although they systematically vary with the functioning of an information-processing mechanism, they take no part in the processing (see Kosslyn and Pomerantz 1977). None of the models of imagery based on artificial intelligence research treats the images that people report experiencing as functional representations (see Baylor 1971; Farley 1974; Moran 1973; Pylyshyn 1973; Simon 1972). Thus, this issue must be resolved before we can even begin to understand imagery.

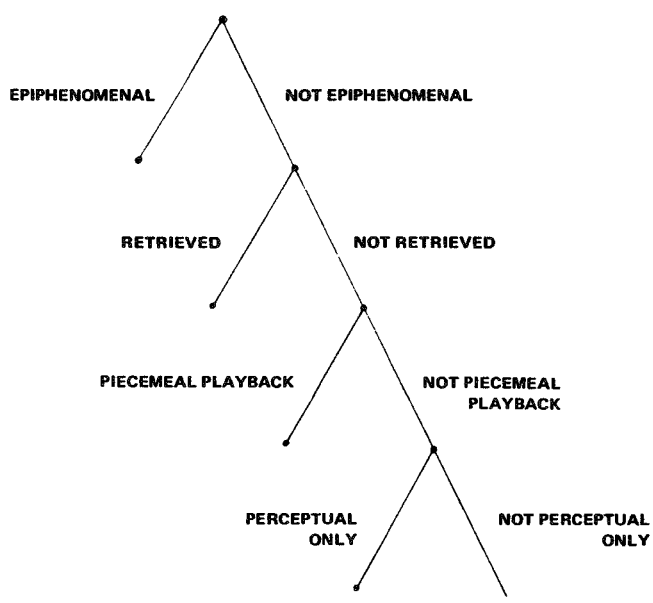

Figure 1. The decision tree at the conclusion of Phase I.

Four classes of experiments were performed to address the image-as-epiphenomenon view. These experiments were motivated primarily by the claim that experienced images "depict" information in a spatial medium (in relation to the interpretive processes that operate on the image). If a representation *depicts* an object, then any part of that representation is a representation of the corresponding part of the object. For example, the rear portion of my *image* of a car is a representation of the rear portion of the *car*. This property is not true of nondepictive representations. For example, "my" is part of "my car," but "my" is not part of the car itself (Ned Block, personal communication, 1979). Because a quasi-pictorial image depicts, it also has the following property: size and orientation of an object *must* be represented whenever a shape is represented; these properties are inextricably linked in the quasi-pictorial format. Thus, if images are in fact functional, then factors like spatial extent – which is inherent in the way visual images depict information – should affect information processing when images are used. In contrast, if our spatial, quasi-pictorial images are not functional, then their spatial properties (which do not characterize listlike linguistically-based representations) would not be expected to affect information processing.

Scanning visual images. If images depict spatial extent, then they should be capable of preserving relative metric distances between portions of objects. If so, then we might expect that more time should be required to scan longer distances across images. Kosslyn (1973) in fact found that the farther a property was from an initial focus point on an imaged object, the longer it took to "see" it in the image. Unfortunately, there was a major flaw in this experiment: more items were scanned over when scanning longer distances (see Lea 1975). Thus, one can explain the apparent effects of distance without referring to spatial images by arguing that all functional internal representations consist of networks of propositions (see Anderson and Bower 1973; Pylyshyn 1973). When subjects (*Ss*) are told to focus on a location on an image, what they really do, the argument goes, is activate a particular portion of a network. When a property is presented, the relevant variable is the number of links in the network that must be traversed before reaching the representation of the property. Because representations of more distant properties are separated from the activated location by more intervening links, more time is required to shift activation to these representations.

Kosslyn, Ball, and Reiser (1978) report a number of experiments that eliminated the confounding between distance and number of items scanned. In one, *Ss* scanned three different distances and scanned over zero, one, or two letters. The number of letters and the distance scanned were varied independently. The time it took to classify the case of the "destination letter" increased linearly with distance and – independently – with the number of letters scanned over. In another experiment, *Ss* first learned to draw a map of a mythical island that contained seven objects (e.g., a hut, a tree, a rock). These objects were located so that each of the 21 interobject distances was at least 0.5 cm longer than the next shortest. After learning to drap the map, *Ss* were asked to image it and to focus mentally on a given object when it was named (each object was used as a focus point equally often). Following this, a probe word was presented; half the time this word named an object on the map, and

half the time it did not. On hearing the word, S was to look for the object on his image. If it was present, he was to scan to it and push a button upon arriving at it. If it was not found on the imaged map, he was to push another button.[1] As before, the longer the distance, the more scanning time was needed.

A control was included in the map-scanning experiment to rule out explanations to the effect that some sort of (unspecified) underlying representation was actually being processed, and that the image itself was epiphenomenal. These Ss participated in the same task as the experimental group, but with one change: after imaging the map and focusing on a named object, they were simply to decide – without necessarily referring to the image – whether the probe word named an object on the map. Thus, if the processing of some abstract representation merely associated with the image was actually responsible for the distance effects observed before, we should find the same pattern of results here. In fact, distance had absolutely no effect on judgment time in this situation.

Finally, another experiment reported by Kosslyn, Ball, and Reiser allows us to rule out one more counterexplanation for the scanning results: on this view, (1) the closer two objects or parts are, the more likely they are to be encoded into the same "chunk," and (2) objects in the same chunk are accessed more quickly than those in different chunks. In this experiment, Ss imaged three schematic faces, with eyes three different distances above the mouth. Immediately after the picture was removed, S was asked to image the face at one of three subjective sizes. Interestingly, the time to scan from the mouth to the eyes (and classify their color) increased not only with the amount of separation between the eyes and mouth, but also as subjective size (and overall distance) increased. The effect of subjective size cannot be ascribed to the effects of distance on initial encoding because subjective size was not manipulated until after the picture was encoded and removed.

Taken together, these results seem to indicate that images do represent metric distance and that this property affects real-time processing of images (see also Experiment 5, Kosslyn 1978a).

Imaging to the point of overflow. The notion that images have spatial extent suggests that they also have spatial boundaries (after all, they don't extend indefinitely). If images occur in a structure specialized for representing spatial information (e.g., within a matrix, as in the CRT protomodel), then the maximum spatial extent of an image should be constrained by the extent of the structure. This idea was tested in the following way: Ss were asked to image an object as if it were being seen from very far away, and then to imagine that they were walking toward the object. They were then asked if it appeared to loom larger; all replied that it did. At some point, it was suggested, the images might loom so large as to "overflow." At this point, S was to "stop" in his mental walk and estimate how far away the object seemed to be, either verbally or by moving a tripod apparatus the appropriate distance from a blank wall. The distance estimates and the length of the longest axis (which accounted for the most variance in distance estimates in a regression analysis) of each imaged object were then used to calculate a visual angle subtended by the image at the point of overflow. (This basic experiment was performed in a variety of ways, which differed in terms of how distance was estimated and in terms of whether Ss mentally imaged pictures that were visually presented or animals that were described.)

In all the experiments, the basic results were the same. First, Ss 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). In fact, distance estimates usually increased linearly with the size of the imaged object. Second, the estimated visual angle at the point of overflow generally remained constant for objects of different size when pictures were imaged. (When named animals, not shown just prior to being imaged, were used as stimuli, however, the angle sometimes decreased for the larger ones.) In addition, in another experiment, similar estimates of the "visual angle of the mind's eye" were obtained by (1) measuring the amount of time required to scan across the longest possible nonoverflowing imaged line, and (2) simply asking people to gesture the apparent size, using their hands, and then measuring the spread and distance from the eyes. These results, then, support the claim that the images we experience are spatial entities and that their spatial characteristics have real consequences for some forms of information processing (see Kosslyn 1978a).

Subjectively smaller images are more difficult to examine. The CRT protomodel suggests that images are processed by the same sorts of classificatory procedures as are used in classifying perceptual representations. If so, we might expect some of the same constraints that affect ease of classifying percepts also to affect ease of classifying parts of mental images. An obvious example is apparent size: parts of subjectively smaller objects are harder to perceive visually; thus, we might expect that they are also harder to see in a mental image. Kosslyn (1975) tested this idea in a variety of experiments requiring Ss to imagine animals in different subjective sizes. In all of these "imagery detection experiments" S was told that we were interested in how long it took to see a property on the image or to see that it was not there. Only after the property was either clearly in view or clearly not present was S to respond by pushing the appropriate response button. The results indicated that more time was required to see properties on subjectively smaller images. Ss often reported that it was necessary to "zoom in" on an initially small image to see a property, but that such zooming in was not necessary to examine a larger image.

As in the initial scanning experiment, one might try to explain these findings with a model in which all information is stored in networks of abstract propositions (e.g., see Bower 1978). That is, perhaps the representation of the concept of an animal includes a list of properties of that animal, and the size manipulations merely vary how many of these properties are activated prior to the probe. On this view, people realize that they should "see" more things on a larger image, and thus they activate more entries on the property list when a "larger" image is required. Hence, the probability that a given probed property is activated prior to query is higher when subjects are asked to form subjectively larger images. And verification time is faster if the property sought is already activated than if it must be searched for in long-term memory, as would happen when people are asked to form a subjectively smaller image, thereby initially activating fewer properties (cf. Anderson 1978).

An experiment was conducted to distinguish this notion from the present one, which posits that size per se is important. If the effects of subjective size are simply a consequence of probability of activation on a list, then we would expect Ss to be faster in verifying properties stored near the "top" of the list (because these properties are most likely to have been activated at the time of probe). Theorists have inferred that lists are ordered by association strength or the frequency of cooccurrence between a noun and a property from the fact that highly associated, frequent properties are verified most quickly in standard sentence verification tasks (when Ss are simply asked to decide as quickly as possible whether a given property is characteristic of a given object – see Conrad 1972; Smith, Shoben, and Rips 1974). If lists are so ordered, then the association strength of a property – and not its size – should dictate the time needed to see the property on an image. This idea was tested by constructing items such as "cat claws" and "cat head," where the smaller property was more closely associated with the noun (as determined by normative ratings). Interestingly, people saw these larger properties more quickly when asked to find them on an image of the object. When no imagery instructions were given, people were faster in verifying the smaller, more closely associated properties. The same results were obtained using a very different technique involving a regression analysis on times to evaluate items not selected for the size-association strength tradeoff (see Kosslyn 1976). These results, then, allow us to distinguish between processing of images and nonimaginal representations.

Effects of the subjective size of an image on later recall. It has also been demonstrated that smaller images are remembered less well in

an incidental memory task. This result was found in four experiments, which controlled for the amount of effort required to form the image at different sizes and the relationship between two objects in an image (one of which may have been imaged at a tiny size; see Kosslyn and Alper 1977).

Other findings in the literature. In addition to our own experiments, Cooper and Shepard (1973; 1975), Cooper (1975), Cooper and Podgorny (1976), Bundeson and Larsen (1975), Larsen and Bundeson (1978), Sekuler and Nash (1972), Shepard (1978), and others have shown that images under transformation (e.g., mental rotation, size change) behave like spatial "analogues" of the represented object (see Kosslyn, in press, for detailed reviews). These results on image transformations are not difficult to explain without referring to quasi-pictorial images (see Anderson 1978), but they were predicted by, and are (in our view) most elegantly explained by, theories positing functional quasi-pictorial images.

The total weight of the evidence, then, supports the view that images are not simply epiphenomenal concomitants of more abstract underlying processing. The results are much more simply explained by positing functional, spatial/quasi-pictorial images than by formulating "Rube Goldberg" nonimagery models (which are not only ad hoc, but have failed to have the heuristic value for predicting new results that is provided by the imagery models). Thus we decided to descend the branch that rejects the claim that images are epiphenomenal and proceed to the next issue.

1.2 Issue II: are images simply stored intact in long-term memory and later simply retrieved in toto?

The second node of our decision tree is at the bottom of the branch representing the hypothesis that images are functional. This new node represents the issue of the way in which the images we experience arise. Two branches extend from this node: on the one hand, images could be stored in toto and simply retrieved; on the other, images may not simply be replayed or projected holistically. All of the existing experiments in the literature bearing on image formation (e.g., Beech and Allport 1978; Paivio 1975; Weber and Harnish 1974) used scenes containing multiple objects as stimuli. These results do not tell us whether individual images are simply retrieved or can be constructed from parts.

We claimed earlier that larger images are more quickly examined because more information is apparent (to the "mind's eye" interpretive procedures) on them. If so, then subjectively larger images may require more construction time if images are in fact constructed by elaboration (adding more parts or detail). Kosslyn (1975) found this to be the case; subjectively larger images generally required more time to construct, independent of the actual size of the imaged object (within a relatively narrow range). One could argue, however, that this reflects a "criterion effect," not the effects of adding more parts. That is, since more material is packed into a smaller area, perhaps subjectively smaller images reach some level of brightness sooner than do equivalent subjectively larger images. If an image is considered complete when it has reached a given level of vividness, then this effect could produce the observed effects of subjective size on image formation time – even if images are retrieved all of a piece, with no construction.

From this counterinterpretation, one might expect that as more details are added to a picture, less time should be required to report that an image of it has been formed. If the construction idea is correct, in contrast, more detailed pictures should take more time to image. These contrasting predictions were tested in an experiment in which *Ss* formed images of more or less detailed versions of pictures of animals. *Ss* studied a picture, imaged it (pushing a button when the image was completed, allowing one to measure the time required to evoke the image), answered a question about the image, and then chose which picture (from a set of two) they had imaged. The results indicated that people do, in fact, require more time to image more detailed pictures.

A control group was not asked to use imagery in this task, but was simply asked to answer as quickly as possible. Instead of pushing the button when they had an image, these subjects pushed it when they had "quickly reviewed the properties of the drawing" in their minds. Unlike the imagery task, this review process required the same amount of time for detailed and undetailed pictures. Thus, imagery was distinguished from nonimaginal retrieval, and the view that images are not simply played back or retrieved in toto was supported.

One could argue that more detailed pictures required more time to image not because they were constructed but because there were more things to check after the image was retrieved. That is, one may indeed simply "project" an image, requiring the same amount of time for detailed and undetailed ones. But after the image is present, one may first check over it before deciding that the image is in fact fully retrieved (even retrieval of a single unit need not be instantaneous). And because one scans to more parts on more detailed images, more time is consumed before deciding to push the button. This idea was examined in a number of experiments (see Kosslyn, Reiser, and Farah, submitted). In one, *Ss* were asked to image sets of objects at different distances from each other. Image formation time increased with the number of objects, but not with the distance – although time to scan between pairs after the image was formed did increase with distance (replicating the earlier results). If the effects of number of objects were due to scanning, distance should have affected times here. Further, in another experiment the differences in time to form images at different subjective sizes were eliminated when simple line drawings were used as stimuli – parts of which were equally easily "seen" (and hence presumably equally easily inserted into the image during construction) at the two sizes. In this case, if scanning were at the root of the complexity effect, we would again have expected less time to scan over and examine the smaller images.

1.3 Issue III: are images retrieved in units or piecemeal?

Given that images are not merely turned on like slides in a projector, are they retrieved in coherent units or simply retrieved piecemeal, in no particular organized fashion? People may have integral representations in memory that are sampled (perhaps at random places) and activated a portion at a time. Alternatively, coherent units could be retrieved and composed in the act of construction. This question was initially addressed in an experiment wherein people imaged drawings of matrices of letters. The letters used to compose a matrix were arranged to form greater or lesser numbers of units according to Gestalt principles of organization (e.g., six columns of six letters were evenly spaced or grouped to form two wider columns, three letters across). More time was, in fact, required to image drawings containing more units (even though the same number of elements was present).

One interpretation of these results would be that the image is stored integrally but that the retrieval process segregates the representation into units during construction. To consider this possibility, another experiment was performed. In this experiment units were defined by presenting parts of a stimulus separately over time. Thus, even though the final products occupied the same area and had the same number of lines, the number of units was varied by varying how the stimulus was broken up into parts initially. In this experiment, then, people learned to image a set of drawings prior to the experiment proper. A given drawing was presented in one of three groups (but counterbalancing resulted in each stimulus occurring equally often in each presentation condition), defined by whether (1) an animal was drawn completely on one page, (2) parts were separated and presented on two separate pages (in the correct relative locations), or (3) parts were separated and presented on five separate pages (in correct relative locations). When parts were distributed on more than one page, *S* was told to study each page as long as he liked and to "glue" the parts together in his mind to form the whole animal. This forced *Ss* to encode separate units and to

integrate them in memory. If images were later constructed by composing these units, we would expect *Ss* to require more time to construct images of animals presented on more pages. This, in fact, was the case. Interestingly, however, no more time was required to "see" parts of imaged drawings once the images were formed; thus, it was not simply that *Ss* were more confused about the appearance of drawings that were presented on multiple pages.

The foregoing experiments seem to demonstrate that physical properties of the stimulus can influence how many units are encoded into memory and later used to form mental images. In addition to such "bottom up" procedures, it seemed likely that "top down" conceptual processes could influence how a stimulus is parsed into units and later imaged. Another experiment using a between-subjects design was conducted to investigate this idea. In this experiment, *Ss* in one group were unaware of the existence of the other, and the number of units in a figure was varied between the two groups. *Ss* were shown a set of geometric figures, each of which could be described in two different ways: as sets of overlapping forms or as sets of smaller adjacent forms. For example, one figure could be described either as "two overlapping rectangles" or "a central square with four squares attached." One group of *Ss* received the set of descriptions using overlapping forms and one received the set using adjacent forms. The time required to image the figures later was indeed dictated by the number of units in the description. In fact, image formation time increased linearly with the number of units in the description, even though different *Ss* contributed to the different data points.

Given the results of the foregoing experiments, then, we have good reason to posit that the imagery system has the capacity to retrieve and integrate "chunks" stored separately in memory. Thus, we are justified in concluding that a theory of imagery must explain how images are constructed from organized units stored in long-term memory.

1.4 Issue IV: are images generated only from "depictive" information?

The fourth level in our decision tree also concerns the origins of images. Images could be generated by simply composing "perceptual" units (encodings of "seeing as"), or image construction could involve an interplay between depictive and descriptive memories. A number of experiments were conducted to investigate this issue. In one (see Kosslyn 1978b) *Ss* first viewed a three by six array of letters. After the array was removed, they were told it would be referred to either as "six columns of three" or as "three rows of six." Interestingly, when it was described the first way, with more units predicated initially (and in terms of columns rather than rows), more time was later required to image the matrix. Since it was the same matrix in both cases, and the labels were given after the matrix was removed, these results seem to indicate that conceptual information can influence image construction. Kosslyn, Reiser, and Farah (submitted) report another experiment that also makes this point. Here, *Ss* were able to use verbal descriptions to form images of scenes with objects at different distances from each other. Not only did image formation time increase with the number of items in the scene, but the time to scan between items later was determined by the distance between them. Beech and Allport (1978) present further evidence that conceptual information is used in the formation of mental images. In addition, Weber, Kelley, and Little (1972) present some data indicating that actual verbal (not simply abstract discursive) information is sometimes used in imaging sequences of letters of the alphabet (see Kosslyn, in press, for a detailed review of this literature).

It may help the reader gain a better picture of the overall research strategy if we summarize the four points established during Phase I somewhat more abstractly. The first point demonstrated that there exist data that can be promisingly construed in terms of the central feature of the CRT protomodel, the "quasi-pictorial display." Because this issue is at the heart of the protomodel, data that make it plausible lend credence to the protomodel itself as a conceptualization of the imagery system. The systematicity in the initially collected data was especially important because in the absence of a reasonably defined range of diverse, yet well-behaved data, the task of theory construction is hopeless. At the close of research on the first issue, then, we had data that fit together and complemented one another in an appropriate way, *if* construed in terms of the CRT protomodel. Thus, we felt comfortable in letting the protomodel generate additional questions. The second point followed from an investigation of a question left open by the CRT protomodel, namely, how the information underlying the images we experience is represented in, and later retrieved from, long-term memory. Given the data, it seemed reasonable to conclude that images are not stored holistically in long-term memory and simply "activated" when one experienced a surface image. Hence, the study of image processing will have to answer questions about the nature of the underlying representation and the way it is mapped into a surface image. The third point established was that the image system is built to allow one to construct images from organized units stored in long-term memory. Hence, a process model will have to include provision for combining units to form a surface image. The final point established during Phase I was that image construction can exploit nonpictorial as well as pictorial information from long-term memory, which also must be explained by a theory of the mental image processing system.

Thus, at the end of Phase I there are a number of constraints on the form of a model and a body of data in need of explanation. The model will include a CRT-like display medium, techniques for forming an image on this display, and techniques for interpreting ("seeing") and transforming information in such a display. What requires explanation, then, is how the image is formed from information in long-term memory and how, once formed, the image is used in various cognitive tasks. At this point it made sense to begin to formulate a theory and model.

2.0 An overview of the theory and model

The following brief overview touches on the central aspects of our theory, model, and approach to theory formation.

2.1 The desired form of a psychological theory: theories and models

The present model is grounded in a preconception about the form of an adequate psychological theory. On our view, a "psychological" (as opposed to, say, a physiological) theory ought to specify the "functional capacities" of the brain, the various kinds of things the brain can *do* during the course of cognition. This does not require any direct reference to the brain itself, any more than specification of the functioning of a computer while it is executing a given program requires one to discuss the operation of the hardware. These functional capacities will be of two kinds: (1) the data structures that can occur in memory, which will be specified in terms of their format, organization, and the kind of content that can be stored; (2) the operations that can process these data structures, which will either transform data structures in some specifiable way over time or will interpret data structures (including comparing two data structures or parts thereof). The theory should not only specify the operations, but should specify the input conditions required by each (including availability of particular data structures) and the results of having executed a given operation. The input conditions and output characteristics of each process result in "rules of combination," which constrain the order in which given operations can be executed in sequence. A complete cognitive theory, then, would allow one to explain performance in all the tasks in the domain of the theory. That is, the theory will allow one to specify the ordered sequence in which operations process particular data structures when people perform a given task. The number of such operations, their individual complexity, and so on will allow one to account for the amount of time

necessary to perform the task, the probability of errors being committed when one is performing the task, and so on.

The reader should note that the actual expression of the theory may not preserve the individual "functional capacities" (data structures and processes) as distinct terms. That is, it may turn out that a more perspicuous statement of the theory can be made mathematically by grouping various capacities together at more abstract levels. We make no commitment as to the particular form of such an ultimate abstract expression, but only claim that it will express lawful relations among the kinds of cognitive entities described here. Thus, our job at this time, as we see it, is to isolate and describe the individual functional capacities and to try to develop the most perspicuous statement of these capacities.

The problem, then, is how to begin to formulate a theory of the sort outlined above. One way to begin is to develop a *model* of the presumed functional capacities. A model, as the term is being used here, is conceptualized as a "range" into which the properties of a given "domain" (in this case, image processing in the brain) are being mapped. The theory picks out the relevant properties of the domain under investigation and maps those into selected aspects of the model. Thus, the relation between the model and the modeled domain is one of analogy. Under the correct description, the model captures the theory-relevant properties of the domain of study. Thus, a model airplane in a wind tunnel is a model insofar as the shape of the wings, fuselage, and so on accurately reflects that of the plane in question; the fact that the weight, type of material, color, and so on are not the same in the model and the actual airplane are of no consequence. The assignment of what is taken to be an "important" and an "unimportant" aspect of the model is made on the basis of the theory, which tells one what is important about the plane itself. Thus, one way to test a theory is to construct a model, which will be a particular instantiation of (at least some of) the system of lawful relations expressed in the theory. This technique is especially useful when one does not have a complete theory, forcing one to add properties to the model simply to obtain accounts of data; in the present case, in a cognitive model these properties will be new functional capacities or properties thereof. If these properties turn out to be important, this provides an impetus to study them in their own right. In addition, it may turn out that a property of the model previously regarded as incidental (e.g., in the case of the model plane, the material) is in fact important. This would provide motivation for further development of the theory itself. Further, when building the model, one may discover alternative ways of instantiating some principle and thus be led to perform experiments to eliminate possible alternatives – again resulting in further elaboration of the theory. In our example, if the amount of sweep of the wings is not specified initially, one will be at an impasse when building the model. In such a case, one would proceed to examine the domain itself, the plane, to discover what the actual shape is and how it should be modeled.

It is useful to distinguish between *specific* and *general* models. Specific models are designed to account for performance in a particular task, whereas general ones embody the set of principles that should account for performance in all the tasks in a given domain. The problem with specific models is that it is difficult to be sure that any theoretical claims that emerge from developing them will be consistent with claims derived from other models. In a general model, since all the proposed functional capacities are available to be used in performing *any* task, one is forced to define precisely the input conditions and output characteristics of each process, and to be consistent across tasks. Use of a general model seems to ensure that a real accumulation of underlying regularities can occur, that functional capacities can come to supplement and complement each other. We have embodied our general model of imagery in a computer simulation. In our simulation, each process is represented by a distinct procedure, and each data structure has been modeled as well. The rules of combination are implicit in the conditions that call upon a procedure and the kinds of suboperations (including the specification of which data structures can be accessed in a given situation) that are permitted by each procedure. Further, given a particular input configuration, if we specify the rules of combination precisely enough, only one sequence of operations will be evoked – providing a specific model of how a particular task is accomplished. Thus, we assume that although logically a given task *could* be accomplished in more than one way, and in fact may be done differently on different occasions (e.g., when tired or well rested), on any *given* occasion the total input configuration and state of the system at the time will uniquely determine the way a task is performed.

In any case, we continue to distinguish between the theory itself, which is abstract, and our model at a given time, only some of whose features will be motivated by the theory. The major purpose of the model is to force us to continue to elaborate the theory – both by introducing new functional capacities and properties thereof and by continuing to define more rigorously which features of the model are theory-relevant and which are not.

2.2 The simulation technique

There are at least five reasons for constructing a computer simulation model; first, it forces one to be explicit. Second, it allows one to be general and detailed at the same time: in a way, the program serves the function of a note pad in arithmetic, saving one the effort of keeping too many things in mind at once. Third, there is a level of abstraction at which features of computational models correspond closely to features of a cognitive theory. That is, the simulation medium allows us to embody our cognitive theory in a general model of the sort discussed above. Hence, the fate of certain aspects of the model can have direct implications for the development of the theory itself. Fourth, it allows one to know whether one's ideas are sufficient in principle to account for the data. If the program runs as expected, it is a kind of sufficiency proof. Fifth, computer simulation helps one to make predictions on the basis of complicated interactions among components, many of which were initially included in the model for entirely different reasons. This last virtue of the simulation technique becomes increasingly important and interesting as a model and theory become relatively detailed and precise. In the present paper we have chosen to focus on the process of using a model to help one develop a theory – which has been the primary use to which we have put our model until relatively recently. Nevertheless, one should not underestimate the importance of this last virtue, and the interested reader is referred to Kosslyn (in press) for a detailed discussion of the predictive utility of the present simulation model.

We wish to emphasize that the present approach does not simply consist of constructing a detailed model and then testing it; rather, we attempt to develop the model – and theory – by performing experiments to help discriminate between alternative implementations. Constructing the simulation helps us discover new, overlooked, or unfathomed issues and questions. In addition, as the theory begins to take shape, actually running the simulation helps one discover the implications of the theory, which can then be tested directly.

2.3 The theory and general model

In the following section we will first review the data structures posited by the theory, describing how these have been instantiated in the model. Following this, we will consider the basic classes of procedures posited thus far, describing how they are concatenated to model the generation, inspection, and transformation of visual images.

2.3.1 Data structures

The simulation contains two sorts of data structures: a "surface matrix," representing the image itself, and "long-term memory files," representing the information used in generating images.

The image proper. The actual "image" data structure that depicts information is represented by a configuration of points in a matrix. The matrix corresponds to a "visual buffer," which is posited to be a spatial medium also used to support the representations underlying the experience of seeing during perception. A quasi-pictorial image is displayed by selectively filling in cells of this matrix. This "surface" image was designed to simulate five properties of imagery discovered in the experiments described above.

First, the image depicts information about spatial extent; it is also capable of depicting information about brightness, contrast, and the like. Thus, parts of the surface image correspond to parts of the represented thing, and the interpoint spatial relations among the thing's parts are preserved in the image. These properties were implied by the results of the experiments described in Section 1.1

Second, the degree of activation (i.e., the maximum contrast between filled and unfilled cells) decreases with distance from the center until no cells are activated (material in this region has "overflowed"). This property was suggested by the finding that the estimated absolute size of the mind's eye (measured in Kosslyn 1978b) was reduced when a stringent definition of overflow was provided in the instructions. This seems to indicate that images gradually fade off toward the periphery, that overflow is not all-or-none.

Third, the surface display has limited resolution, causing contours to become obscured if an object is pictured too small (because there are not enough cells per unit area to depict details). This property is based on the finding that subjectively smaller images are more difficult to "inspect" (see Kosslyn 1974, 1975, 1976; Kosslyn and Alper, 1977).

Fourth, the spatial medium within which images occur has only a limited extent, and it has a definite shape. The data reported by Kosslyn (1978a) and Finke and Kosslyn (in press) indicate that the most activated region of this display is roughly round in shape, but that the shape flattens into an ellipse in the less activated regions.

Finally, the matrix corresponds to a "visual short-term memory" structure. Representations within this structure are transient, requiring effort to maintain. That is, as soon as an image is placed in the matrix, it begins to fade, and effort is required to "refresh" portions of the image to prevent it from disappearing. In fact, if an image is very complex, initially refreshed portions may have faded altogether before other portions have been refreshed, defining a "maximum complexity capacity" of the buffer. This property was motivated by results reported by Kosslyn (1975) indicating that more complex images are more difficult to maintain.

Long-term memory representations. There are two sorts of representations used in the generation of images, one storing information about the "literal" appearance of an object, and another storing facts about the object.

The perceptual "literal" memory of the appearance of an object or scene is not interpreted semantically; it is the product of "seeing that," not "seeing as." This memory of how something looked is what allows reproduction of the experience of "seeing" during imagination. In the model, this information specifies how points should be arranged in the surface matrix to depict some object. This information is represented with r, Θ (radius, angle) polar coordinate pairs. This representation was chosen because it allows images to be generated easily at different locations in the surface matrix and at different subjective sizes (i.e., sizes in the matrix). In earlier experiments (Kosslyn 1975; Kosslyn, Reiser, and Farah, submitted) it was found that people can, in fact, form images at different sizes directly and can "project" imaged objects at different relative locations directly. In addition, although we did not select the format with this in mind, this format implied that images could be generated at nonstandard orientations (by altering the Θ values); we have been investigating this "accidental" feature of the model and now have preliminary data supporting this prediction, making this property desirable. This, then, is an example of the model's heuristic value – once we had decided on a tentative representational format, the format suggested the existence of a particular imaging ability, which subsequent behavioral research seems to confirm.

A given object may be represented by more than a single image file in the model. One such file represents a global shape or central part (this question is open at present). Other files represent "second looks" that can be integrated into the global or central shape to form a fully fleshed-out image. Thus, the theory posits hierarchical organization in long-term memory. This multiple file representation was motivated by the finding that more time was required to generate images containing more units, even if the amount of material (e.g., letters in a matrix) was the same. Pilot data indicated that people could construct images of pictures of different complexity in about the same amount of time if they were not required to include details. Some *Ss* reported that they formed somewhat fuzzy images of more complex objects and then waited until these were probed before filling in the details. This sort of introspection is consistent with the notion that a global image is initially generated and then elaborated, but we have as yet no convincing evidence on this issue.

The second type of long-term representation we posit is a set of facts about the imaged object. These facts are represented discursively, in a "propositional" format, and represent (1) where and how a part (which is itself in a file of r, Θ coordinates) is attached to a "foundation part" of a global or skeletal image, and (2) how to locate the part itself in an image. This information is represented as an ordered set of procedures that will search the "visual buffer" (the matrix, in the model) testing for specific patterns of points (e.g., horizontal lines that intersect vertical ones, etc.). If the set of procedures that describes a part can be successfully executed, the part will have been located. Also represented is (3) the name of the file that contains the "literal" memory of the object or part (i.e., a stored list of r, Θ, pairs), (4) a classification of the object's size, and (5) the name of the most highly associated superordinate category (which is used in answering questions about objects and their parts).

It was necessary to include the foregoing sorts of information in the theory in order to provide models for the ways in which we have found that descriptive information could be used in generating images. In addition, numerous problems were discovered in the course of implementing the model (e.g., if a part is not visible, should one "zoom in" or "pan back" before trying again?), many of which necessitated including particular sorts of information (e.g., the size tag, in this example).

2.3.2 Image processes

There are three general sorts of processes used in imagery: routines for generating surface images, routines for classifying these images or parts thereof, and routines for transforming images.

Image generation. In the model, images are generated by printing out points in the surface matrix, depicting first a "skeletal" image to which details may be added. Three procedures, called PICTURE, PUT, and FIND, perform the computations that generate the images. We claim that people have operations that accomplish the same ends as these procedures, although obviously the human operations are not identical to their counterparts in the simulation.

PICTURE converts an underlying literal memory file into a configuration of points in the visual buffer (the matrix in the model). As input PICTURE takes specifications of size, location, and orientation; in the model, this procedure adds to or multiplies the r, Θ coordinates appropriately before printing the points in the surface matrix. (If no particular specifications are given, default values – which we have determined empirically – are used.) PUT is a procedure used to integrate parts into an image. For each part, it accepts a location (i.e., the "foundation part") and the particular spatial relation between the part and its foundation part (e.g., for car, when inserting an image of the front tire onto an image of the body, the location is "front wheelwell" and the relation is "in and under"). PUT then looks up a set of parameters dictating how the particular spatial relation between the part and its foundation will be translated

into operations on the surface matrix itself. But before one can attempt to attach a part to its foundation, one must *find* the foundation in the image. This is because images can be formed at different subjective sizes and relative locations (Kosslyn 1975; Kosslyn, Reiser, and Farah, submitted), and parts must be integrated via "inspection" of the existing image for the location of the relevant part – an a priori absolute location will not suffice. The FIND procedure takes as input a description of a part, which in our model consists of the list of subprocedures that search for patterns of points. FIND executes each subprocedure in sequence, locating the part if and only if all of the subprocedures can be executed successfully. The PICTURE, PUT, and FIND procedures are coordinated by an executive procedure called IMAGE, which governs whether or not parts will be inserted (in response to various task demands).

Let us consider a concrete example. In generating a detailed image of a chair, the IMAGE procedure first constructs – via PICTURE – a skeletal image of a chair and then searches the "factual" information stored in long-term memory for names of parts that go on a chair. In the current simulation, the fact that chairs have cushions is found, and further the fact that cushions are FLUSHON seats (the foundation part) is also found. PUT then calls FIND to locate the relevant foundation part of the image (the seat) by means of a set of procedures describing SEAT (retrieved by looking up the representation of seat and locating the procedural description). Once FIND locates the foundation part (the seat, in this case), it passes back the Cartesian coordinates of the part's location in the image. PUT then checks the location relation and determines the necessary "subjective" size (i.e., relative size in the matrix) of the part, in this case setting the size of the cushion so it fits flush on the seat. Once the correct location and size are computed, PICTURE is called by PUT, and the part is integrated into the image.

This model allows us to explain why more time is required to form images composed of more parts and leads us to predict a linear increase in image formation time with increased numbers of units (because each unit requires an increment of time to generate). This prediction is in fact generally borne out by the data (Kosslyn, in press). Further, if FIND fails to locate the foundation part (e.g., seat) during image construction, perhaps because the image is too large or too small, the part (cushion) is simply omitted from the image – explaining why subjectively smaller images are generated more quickly but are more difficult to examine (Kosslyn 1975; Kosslyn, Reiser, and Farah, submitted). In fact, the model allows us to provide relatively perspicuous accounts of the available data in the literature on image generation (see Chapter VI of Kosslyn, in press). Among the questions the model leads us to ask about image generation are the following: what are the principles that delineate a "part"? What are the principles that order the parts during generation? When will a detailed image (with parts) be generated, and when will only a skeleton be formed? What image is formed when only a class name is given (e.g., "dog")? A given particular? A prototype?

Image classification. The primary procedure used in image inspection is FIND. However, if a part is not immediately "visible," an executive procedure, LOOKFOR, will call up ZOOM, PAN, SCAN, or ROTATE (described below) in order to adjust the image appropriately. In addition, LOOKFOR may call up the three procedures coordinated by IMAGE (i.e. FIND, PICTURE, and PUT) if a certain region needs to be inserted or elaborated further before a sought-after pattern can be "seen" (i.e., detected by FIND).

Thus, if the task is to decide whether an imaged object has a certain property, FIND works in conjunction with various image transformations in an all-out attempt to locate the sought-after part. Before trying to locate a part (i.e., before attempting to discover whether the procedural tests describing it can be satisfied), LOOKFOR looks up the relative size of the part and calculates the resolution of the image (i.e., the dot density, in our model) that would be optimal to "see" the part. If the dot density is not within the optimal range, it calls up the ZOOM procedure (actually, ZOOM or the inverse, PAN) to expand or contract the image (as appropriate) until the resolution *is* optimal. As an image is expanded, the relevant region likely to contain the sought-after part is further elaborated as it becomes possible to insert new portions into the image (because more and more foundation parts become discernible – see Chapter V of Kosslyn, in press, for a more detailed description of this process). If the size of the image is correct, and the correct region is in focus, but the classificatory procedures executed by FIND are still not satisfied (e.g., it cannot locate the configuration of points delineating the arms of a chair), the program returns a failure message. If it locates the part, it returns an affirmative response. The present model allows us to account for the basic data on image inspection (e.g., more time is required to see parts of subjectively smaller images), as is developed in detail in Chapter VII of Kosslyn (in press). The following are some of the questions raised here: Under what conditions must a part be filled in only when needed? When is it likely to be "on" the image at the time of a probe? What is the resolution of the visual buffer, and is this the sole constraint on the ease of "seeing" parts of images? How much of an image can truly be "seen" simultaneously, without requiring scanning?

Image transformation. In the simulation, four basic imagery transformations are modeled: zooming in on a part of an image, panning back to "see" more global aspects of an image, rotation of images, and scanning across images (i.e., changing the point of focus). ZOOM consists of moving the points depicting an object outward from the center of the image, beginning with the outermost points. PAN involves pulling points toward the center, beginning with the innermost points. ROTATE involves shifting the points around a pivot. Scanning an image (accomplished by SCAN) is at present treated as another kind of transformation, in which points are moved across the surface matrix so that different portions of the image seem to move under the center (which is most highly activated and in sharpest focus).

The single most important result in the mental image transformation literature that our simulation must account for is that transformation time is proportional to the distance or amount of transformation. This result has been found in experiments investigating size change (Sekuler and Nash 1972; Bundeson and Larsen 1975; Kosslyn and Shwartz 1977), scanning (Kosslyn 1973; Kosslyn, Ball, and Reiser 1978), and rotation (Shepard and Metzler 1971; Cooper and Shepard 1973; Cooper and Podgorny 1976). In the present theory, the general interpretation of these findings is that mental images are transformed in small steps, so the images pass through intermediate stages of transformation during the transformation process. Hence, the greater the degree of transformation, the more increments are necessary – resulting in increasingly more time being needed to accomplish the transformation.

In the simulation, transformation operators work directly on the surface representation by "moving" the dots that depict the imaged object. The image is transformed incrementally for a number of reasons. Typically, an image is transformed because the resolution, point of focus, or orientation is not optimal for a particular inspection process. Transforming the image gradually allows constant checking to determine whether the necessary resolution, point of focus, or orientation has been reached. An "all at once" transformation, in which the transformation is done in a single step, has several difficulties: first, one must calculate *exactly* how much to transform the image. A single large step, if overestimated, would cause the transformation to overshoot the desired resolution, in which case several additional small steps would be required anyway. Second, the transformation operators are unlikely to be perfect. If any distortion is introduced into the image by the transformation operator, it is probable that the degree of distortion will be proportional to the size of the transformation step. Simple cleanup operators (e.g., contour sharpeners) could be applied to the image after each small step, whereas more complex cleanup operators (e.g., ones requiring knowledge of the details of the pattern) would be required after a large step. Finally, small step transformations allow a simple control structure – each iteration is the same. A transformation process that takes a big step as a first approximation would require a more

complicated control structure, so as to allow one to shift gears in the middle of the transformation. For example, not only might the large step have to be followed by several small ones, but various classes of cleanup processes might have to be invoked after the different step sizes, and provisions would have to exist for reversing the direction of the transformation if an overshoot occurs. The main point here again is that the model provides a framework for asking questions whose answers will accumulate in articulating the theory.

Up to this point, we have considered only image transformation procedures that work directly on the surface representation. The data structures of the simulation, however, imply that there may be another, fundamentally different mode of transforming images: since the simulation allows for images to be generated initially at any size, location, or orientation, it follows that an image can be transformed simply by "erasing" the image in the visual buffer (or allowing it to fade) and generating a new image at the correct size, location, or orientation. We will refer to this mode of imagery transformation as a "blink" transformation, in contrast to the "shift" transformations that occur incrementally.

As mentioned above, research on imagery transformations finds that, left to their own devices, *Ss* typically use shift transformations. That is, transformation time is usually found to be proportional to the distance or amount an image is transformed, whereas the amount of time to perform a blink transformation should be independent of distance or amount per se. We hypothesize that the preference for shift transformations results from an increase in the amount of effort or time necessary for working from the deep representation, and we are currently testing this notion. Our results so far indicate that when *Ss* are instructed to perform a blink size transformation, they are in fact generally slower than when instructed to perform a shift transformation. Results on scanning images indicate that shift scans are faster up to a certain distance, after which blink scans are faster. This latter result makes sense when one considers that shift transformation time is proportional to distance, which blink transformation time is not, and thus there may be some distance at which shift transformations become slower than blink transformations.

We have found it useful to introduce yet another distinction pertaining to mental image transformations. In our model, there are two general classes of transformations (which can be performed either by shift or blink operations). "Field-general" (FG) transformations operate by uniformly translating all the contents of the visual buffer (the matrix in our model) in some way; "region-bounded" (RB) transformations delineate a particular region in the visual buffer and only transform the points within it. Virtually every FG transformation has an RB analogue. For example, "zooming" is FG, but "growth" (i.e., imagining the object growing larger – not moving closer) is RB; reorientation (i.e., imagining tilting your head) is FG, while rotation is RB; scanning is FG (where the image is shifted so that different portions fall under the center of the visual buffer, which is most highly in focus), whereas position translation (e.g., imaging a saltshaker sliding along a tabletop) is RB.

In general, we expect to find effects due to the complexity of an imaged scene only with RB transformations, since it should be more difficult to delineate a to-be-transformed region in a complex environment. For example, the more objects that must be held in an image while one object is being manipulated, the more time should be taken. Since FG transformations operate uniformly on every point in the surface display, the actual contents of the image should be irrelevant, and we would not expect any effects due to, say, the number of objects in an imaged scene. Thus, if one images zooming in on a scene, the number of items in the scene should not affect time; but if one images the items actually growing, more time should be required when more items are included in an image. Similarly, no more time should be required to scan a scene with more items, providing the same number of items is scanned over in all cases. This was in fact found to be true (see Pinker and Kosslyn 1978). But the time required to move only one part in relation to the others should increase with more complex objects or a greater number of objects in the scene, a prediction that is consistent with findings reported in the recent Pinker and Kosslyn paper (although at that time we had not yet drawn the FG/RB distinction and had no way of accounting for the finding). Thus, the model has been especially rich in introducing new distinctions and hypotheses about image transformations. If the data warrant it, many of these properties of the model will become properties of the theory itself.

3.0 The remystification of mental imagery: objections and replies

The program of research we have described has addressed itself to the two classical objections to the use of mental imagery as an explanatory construct in psychology (see Ryle 1949; Pylyshyn 1973). The first objection was that the notion of a "mental image" is either intolerably vague or logically incoherent. We feel that the computer simulation of imagery and the theory embodied in it should lay this objection to rest. The second objection was that mental imagery, however defined, is not an autonomous component of the human mind, but that the processes underlying images are continuous with those underlying more abstract forms of thought. We have two answers to this objection: first, we described earlier a number of empirical results demonstrating that imaginal and nonimaginal thought processes have different operating characteristics. Second, the mere fact that a theory of imagery per se could have explanatory value and predictive utility leaves us loath to abandon such a theory. If in fact imagery is just a special aspect of more general processes, it is surprising that one should be able to make progress in studying imagery in its own right, without focusing on nonimaginal processing per se.

Recently, however, there have been new objections to the study of imagery as it is embodied in this and other research programs. In this section we reply to these objections and attempt to reaffirm the soundness of our approach. The first objection is that our experimental results do not speak to the properties of imagery because of contamination from "demand characteristics" and experimenter effects. The second objection is that even if the experimental results are valid, attempting to distinguish among theories of the representations underlying imagery is misguided, because the task is impossible *in principle*. The third objection is that even if one can distinguish among theories in principle, theories of the sort we have proposed are not explanatory in the way that cognitive theories should be. Finally, we will consider some questions about how one should evaluate our model and its roles in the research program.

3.1 Are subjects just doing as they're told?

Martin Orne (1962) has alerted psychologists to the danger that many of their experimental results can be attributed to the "demand characteristics" of the experimental setting. That is, *Ss* may deduce the purpose of the experiment in which they are participating and may manipulate their responses so as to give the experimenter (*E*) the results they think he wants [see Rosenthal & Rubin: "Interpersonal Expectancy Effects" *BBS* 1(3) 1978]. We are occasionally criticized on these grounds for two reasons: first, we often explicitly instruct *Ss* to use imagery in performing some task (the rationale for this is simple: if we want to use data to make inferences about imagery, we must first be confident that the data do in fact reflect image processing). Second, *S* is often instructed to respond only after a certain condition has been met in an image, a condition that is detectable to *S* and *S* alone. For example, *Ss* in our experiments are asked to respond when they have "seen" a particular object in an image, or when they have transformed it in some way.[2]

Consider the image scanning experiments of Kosslyn, Ball, and Reiser (1978). A critic could argue that *Ss* in this setting have good reason to believe that it will take longer to scan longer distances and will ensure that their responses fit that pattern in order to be "good subjects." Thus, response times will vary linearly with the distance between objects simply because *Ss* wait proportionally longer before responding to targets separated by greater distances,

without in fact scanning an image. Although one could raise similar objections against other imagery experiments, we will concentrate on the criticism as it applies to mental image scanning experiments in particular for two reasons. First, these are the experiments that have been denounced most explicitly (Richman, Mitchell, and Reznick, 1979). Second, we have now begun to use image scanning as a "tape measure" for various geometric properties of images (Kosslyn 1978b; Pinker and Finke, in press; Pinker and Kosslyn 1978), an enterprise that depends on the validity of the phenomenon of image scanning itself. In this section, we examine specific versions of the "demand effects" criticism as it could apply to image scanning experiments and attempt to refute each in turn.

One possible objection is that our experiments allow many opportunities for *E* to influence *S* by nonverbal cues, tacit messages between the lines of our instructions, loaded answers to questions, and so on (see Rosenthal and Rosnow 1969). On this view, we can influence *Ss* to give us *any* results we want; in fact, if our thinking had led us to expect *longer* scan times for *shorter* distances, we would have been able to obtain those results. Of course, precautions are always taken to reduce these possibilities, such as prearranged instruction "scripts," rigid criteria for success in learning tasks, and so on. But the best refutation of this particular criticism is the abundance of instances in which our *Ss* surprise us by responding contrary to our expectations. For example, in Kosslyn, Ball, and Reiser (1978), it was initially predicted that distance effects on scan times would evaporate when *Ss* were instructed to zoom in on one object, keeping the rest of the image "out of view of their mind's eyes." We were wrong. In Pinker and Kosslyn (1978), *Ss* had to scan images containing four or six objects, and occasionally had to "move" one object in an image to a new location. Among our predictions were these two: the rate of scanning would be slower for the image with six objects, and the time to move an object would be the same for images with four and with six objects. We were wrong on both predictions. Anyone who doubts our sincerity on this matter is welcome to read the Pinker and Kosslyn paper, in which we obviously flounder around trying to account for the latter finding. (That was before we had posited, for entirely independent reasons, a distinction between "region-bounded" and "field-general" image transformations, which turned out to be consistent with the discrepancies in that experiment.)

A second version of the criticism is that our scanning instructions connote physical movement, and that *Ss* mentally simulate motion with the result that longer distances are "traveled" in longer times. This criticism has been raised implicitly by Richman, Mitchell, and Reznick (1979) in the title of their paper, and in their concentration on the particular scanning experiment in which a map of an island was used. It is easy to show that the mention of physical motion is *not* a necessary condition for distance effects in scanning; in Kosslyn (1973) the instruction was simply to "shift attention" across an imaged boat; in Experiment 4 of Kosslyn, Ball, and Reiser (1978) it was to "glance up" from the mouth to the eyes of an imaged face. Further, Spoehr and Williams (1978) have shown that longer distances in images take longer to scan, even if the task instructions do not mention "scanning an image." They had *Ss* decide whether or not three landmarks in an imagined map fell along a straight line and found that response times varied linearly with the distance between the objects, implicating a scanning process. Thus, the use of explicit scanning instructions does not seem to be a prerequisite to obtaining distance effects on scan times.

Perhaps, then, it is tacit knowledge of the visual system that leads our *Ss* to expect distance effects. According to this third version of the criticism, *Ss* are accustomed to moving their eyes from one fixation to another and know that it takes longer to rotate their eyes through larger angles. Incidentally, it is not obvious that eye movement times do in fact increase linearly with increasing distance between fixations. Bahill and Stark (1979) point out that in most diagonal saccades, the muscular activities that rotate the eye vertically and horizontally do not begin and end in tandem, but overlap to various extents, yielding trajectories that range from straight diagonal lines to "L" shapes. Nevertheless, longer distances should yield longer eye movement times on the average, which is what we have found (Pinker and Kosslyn, in press). But this effect still cannot account for certain image scanning results. Eye movement times for three-dimensional scenes vary with the *two-dimensional* or angular separation between objects. Nevertheless, when people scan in depth an image of a three-dimensional scene, their scan times vary with the *three-dimensional* separation between objects, contrary to the pattern found for eye movements (Pinker and Kosslyn 1978, in press). Perhaps, then, *Ss* are not cognizant of eye movements per se, but of the general "scanning" process involved in visually exploring a three-dimensional environment. However, when *Ss* are instructed to scan in depth between objects that are visible in front of them, their response times vary linearly and independently with the three- *and* the two-dimensional distances between objects, representing the additive effects of "mental scanning" and eye movements (Pinker and Kosslyn, in press). Thus, in 3D image scanning experiments, *Ss* do *not* simply reproduce the temporal patterns that they would display in analogous perceptual tasks.

A fourth version of the criticism is that *Ss* discern that the independent variable of these experiments is the distance between objects, and that the dependent variable is response time, and they naturally conclude that one must vary with the other (quite independently of any specific consideration of physical motion, eye movements, etc.). To test this possibility, we routinely ask *Ss* upon completion of the experiment to write down what they think is the purpose of the experiment and whether they used any special tricks or strategies. No subject has ever admitted to having deliberately manipulated his response times in more than 10% of the trials in an experiment, and the vast majority deny ever doing it. As to the purpose of the experiment, a very frequent response is something like "God only knows!," but *Ss* often do mention one or more independent or dependent variables. They mention response time and accuracy as dependent measures ("accuracy" referring here to pressing one key when the object to be scanned to on a particular trial is in the image, another key when it is not – a task that we included, in fact, to reduce the salience of response time to *Ss*). As independent measures, they mention the amount of practice or fatigue, the strength of the semantic associations between objects, the direction of scanning, the colors of objects, the similarity of the sounds of their names, their shapes, absolute positions, and so on, as well as the distances between them. (Occasionally a subject will "deduce" that it should take *longer* to scan between objects that are close together, since they are crowded together and hard to see without "straining" or "zooming in.") If attention to distance as an independent variable and response latency as a dependent variable is responsible for distance effects in scanning, one would surely expect *Ss* who named these variables to show higher correlations between their response times and distance than the other *Ss*. One would also expect that when data from these *Ss* are discarded from the aggregate response times, the time-distance correlation would vanish or drop. With this in mind, separate analyses were performed (in Pinker and Kosslyn, in press, and Pinker and Finke, in press) on data from *Ss* who either mentioned distance and response time or who confessed to manipulating their response times on occasion. These *Ss* exhibited individual time-distance correlations indistinguishable from those of the others; when their data are removed from the rest, the correlations between mean response time and distance *increased* more often than they decreased. Thus it is unlikely that attention to distance and time variables is crucial for *Ss* to take longer when scanning greater distances.

Finally, Richman, Mitchell, and Reznick (1979) have performed a pair of experiments that they feel demonstrates the importance of demand effects in image scanning experiments. Following a general suggestion by Orne (1962), they gave a questionnaire to a group of *Ss* describing an image scanning experiment and asked *Ss* how long they thought they would take to scan a short distance, and how long they thought they would take to scan a long distance, were they participating in the real experiment. *Ss* expected that it would take longer to scan the longer distance. The problem with this experiment is that *it* uses demand effects to support its claim about demand

effects in our experiments. When the issue is stated so bluntly to Ss, with two response slots, and one variable differing from one response slot to the other, they had little choice but to make one response larger than the other. This does not imply that they would attend to this variable if they were Ss in our more complicated experiments (in fact, a majority do not), nor that they would manipulate their response times in accordance with this relation (which virtually all Ss deny doing). In fact, it is possible that Ss would agree to *any* question so obviously stated in pseudoexperiments of this sort, regardless of what their behavior would be in the experiment itself. To test this possibility, we described a typical image scanning experiment to 29 Ss, but mentioned that some objects in the image were highly associated semantically (e.g. tree and grass) and that others were not associated (e.g. hut and rock). When Ss were asked to estimate how long their response times would be for these two conditions, their estimates for the low-association pairs were significantly higher than their estimates for the high-association pairs (2.93 versus 1.66 seconds), even though such effects do not obtain in image scanning experiments. In a second pseudoexperiment, we described the same experiment to a new group of 26 Ss, but mentioned this time that objects that are close together in the image (we cited the shortest distance used in the Kosslyn, Ball, and Reiser map experiment) might be hard to "see" distinctly, paraphrasing the hypothesis of one of Kosslyn, Ball, and Reiser's subjects. The estimates of the Ss reflected this suggestion (1.28 versus 2.32 seconds for long and short distances, respectively), even though the *opposite* effects obtain in the real experiments. Thus, Ss certainly *can* respond to demand characteristics in a questionnaire describing an experiment. Whether they do so in image scanning experiments is a completely separate question.

The second experiment that Richman, Mitchell, and Reznick report was also inspired by one of Orne's suggestions. Ss were given a map with three landmarks, linked by two roads. The two roads were equally long, but one had a road sign labeled "80 MILES," another had a road sign labeled "20 MILES." When Ss imagined the map and scanned between objects, they took longer to scan the path labeled with the longer distance. Richman, Mitchell, and Reznick conclude that demand effects may be responsible for response time differences in all image scanning experiments. (One could, for that matter, make an analogous argument that all estimation of line length is based on demand characteristics, given Asch's [1956] demonstration that demand characteristics can alter S's judgments of relative lengths of lines.) We can hardly dispute the claim that it is within people's ability to alter their response times if they are so motivated; that is the reason why our experiments include such precautions as postexperimental questionnaires in the first place. However, the real issue is whether demand characteristics *are* responsible for distance effects in image scanning in the particular experimental situations in which we obtain them. And on this issue, the Richman, Mitchell, and Reznick experiment says nothing. Of course one can engineer a situation in which demand characteristics will affect response times; distance is certainly not the only possible determinant of latencies in psychological experiments. But the conclusion that their result therefore undermines our explanation of distance effects is a non sequitur. Consider an image scanning experiment in which Ss are required to perform a concurrent mental arithmetic task on some trials and not on others. It would not be surprising if Ss took longer when they were doing long division and scanning than when they were just scanning, distance held constant. From such a demonstration, could one conclude that distance effects in conventional scanning experiments reflect nothing more than Ss surreptitiously doing more mental arithemetic for longer distances than for shorter ones? Certainly not. But this seems to be the logic of the Richman, Mitchell, and Reznick experiment.

Whether demand characteristics are responsible for the results of image scanning experiments, or any other imagery experiments, is of course an empirical question. We have not seen convincing evidence or arguments that image scanning effects are contaminated in this way; rather, we think the evidence points to the opposite conclusion. In any case, we are currently planning experiments that will put the question to a more critical experimental test.

3.2 Anderson's argument for agnosticism

Central to our program of research is the belief that it does not suffice simply to build a model that can account for the facts of mental imagery, but that it is imperative to show that our model is "truer," in some sense, than conceivable alternative models. John Anderson (1976, 1978) has argued that this cannot be done – not for mental imagery, not for *any* domain of cognitive science. Unlike the critics discussed in the previous section, Anderson seems to accept our claim that the data we have gathered are determined by the mechanisms that underlie mental imagery. However, he denies that such data (or any data) justify the inference that these mechanisms include a spatial medium or representation. In a nutshell, the formal proof supporting Anderson's argument is a demonstration that one can add an additional operation and its inverse ("cancelling" the effects of the operation) into any theory to produce another theory that will mimic the first. Since this merely proves that one can concoct a less parsimonious alternative to a given theory, it seems totally uninteresting. Thus, we will not discuss it further here and refer the reader instead to Pylyshyn (in press) for a detailed discussion of Anderson's formal arguments (but see also Anderson's reply to Pylyshyn, also to appear in *Psychological Review*).

To many, the real force of Anderson's arguments comes from his *informal* claim that in many or most cases there exist equally plausible, parsimonious, and principled competing models that account for a given set of empirical data. This claim is supported by a single example: the rotation of mental images. Since we believe that so much of Anderson's argument hangs on this example, we will examine it in some detail here, and we will contrast our theory with the one he proposes, in order to argue against his claim that such theories are indistinguishable.

In our computer simulation models an image is represented as a two-dimensional array of dots, and rotating an image consists of incrementally moving the dots in the array around a center axis. In Anderson's proposed alternative, an image is represented by a set of "propositions" describing the parts of the object and their spatial interrelations, and rotating an image consists of incrementing the parameters in the propositions that describe the overall orientation of the object. As we have mentioned, the principal datum that both models must account for is the linear relation between the time people take to rotate an image and the angle through which they rotate it. In both models, this relation is produced by an incremental transformation process, yielding a series of intermediate states, with more states for greater rotations. Thus the question now becomes: is incremental transformation an equally motivated assumption in both theories, or is it integral to one and added on as an afterthought to the other?

First, let us note that there is no reason why an image representation should be transformed *in any way* in a propositional model. In the Cooper and Shepard-type tasks that produce mental rotation, Ss must decide whether a misoriented stimulus letter is normal or mirror reversed, presumably by matching an internal representation of the stimulus against an internal representation of the normal letter. On our "spatial" account, the stored representation is like a template of the letter in upright orientation, so a successful template match requires transforming the representation of the stimulus letter to the upright. On a typical "propositional" account, the orientation of the letter in the stored representation is stated in a proposition distinct from those representing the interrelations of the parts. Since the match procedures interrogate only the latter propositions, why should the orientation parameter be altered at all? In contrast, in a spatial model one *cannot* represent orientation in a format distinct from the representation of shape. And this is our main point: in a quasi-pictorial representation, *orientation and shape are inextricably linked* because they are "in" one and the same representation. Therefore, the ability to match shapes in a template fashion (which is an independently motivated feature of an image processing system; see Cooper and Shepard 1973; Smith and Neilson 1970) requires that orientation be normalized. Furthermore, the accuracy of that match depends on the accuracy of the reorientation transformation – which

in turn, (as we have mentioned) calls for a series of increments rather than a single step, thus yielding a linear increase in rotation time with the angle of rotation. In Anderson's theory, in contrast, inclusion of an incremental transformation operation is motivated solely by the observed finding; and nonincremental transformations could just as easily have been included.

Let us give Anderson the benefit of the doubt and suppose that the orientation parameters must for some reason be normalized prior to match. And let us suppose that the motivation for this is every bit as compelling as the one we offered earlier for the spatial model. Now, is it logically possible to distinguish between the two models? The answer is still clearly "yes." In the present model, rotation is accomplished by repeatedly sweeping through a bounded region in the surface matrix and shifting the positions of points appropriately. The larger the bounded region, the more operations will be required to perform one increment of the transformation, since more cells must be processed. Thus, we expect that more time should be required to rotate subjectively larger images; further, increasingly greater amounts of time should be required to rotate larger images to greater degrees (since each increment will take more time, compounding as more increments are required). This prediction is based on the fact that not only are orientation and shape inextricably linked in an image, but so are size and shape: an image *must* be of *some* size and orientation. Interestingly, data we have collected suggest that the foregoing prediction will be borne out, but this is not critical here. For present purposes, it is enough to note that we have not yet reached the point where interesting competing models are indistinguishable by any empirical test.

3.3 The cognitive penetration of mental imagery

Zenon Pylyshyn (1978, 1979, in press) has recently stated his case against the use of mental imagery as an explanatory construct in a form slightly different from his argument in his 1973 paper. He argues that within a cognitivist framework, the analogue-propositional debate over mental imagery boils down to a disagreement over what are taken as the primitive or elementary information processes underlying imagery. Consider again the rotation of mental images. On the one hand, an operation such as rotation could be considered a primitive, entailing that (1) it owes its operating characteristics to the physiological properties of the neural tissue in which is is instantiated, and (2) its internal operation is unavailable to, and unaffected by, other cognitive processes. In that case, rotation could be considered an "analogue" process. On the other hand, the rotation operation could be composed of a set of subprocesses, in which case its operating characteristics would be determined by the types and arrangement of those subprocesses (which themselves may or may not be analogue according to this definition). To the extent that these subprocesses can be described as symbolic representations, rules, knowledge, or strategies, the rotation process would be considered "propositional." Pylyshyn then argues that there is a relatively simple empirical test for deciding whether or not a given cognitive process is analogue. If the process can be influenced by other cognitive factors – by what the person knows or believes, or by how the person interprets the units being acted upon – then it cannot be a primitive, analogue process but must decompose into parts that can interact with the symbol structures representing the person's knowledge, beliefs, and so forth. The process is said to be *penetrable* by cognitive factors and cannot be considered an analogue process. Whether its subprocesses are to be considered analogue will depend on the outcome of similar tests addressed to each of the subprocesses. Pylyshyn prophesies that the spatial character of the subprocesses would erode as such an investigation is carried further and further.

Pylyshyn (1979) reports a pair of experiments designed to show that so-called "holistic" mental rotation is penetrable by cognitive factors and hence is not an analogue process. According to the "holistic rotation" theory that he attributes to Shepard and Metzler (1971) and Cooper and Shepard (1973), mental rotation of a figure is a primitive unarticulated process that should not depend on the complexity of the figure or on the nature of the comparison task that follows rotation. In Pylyshyn's experiments, *Ss* are instructed to rotate various geometric shapes into the same orientation as a smaller probe stimulus, and to decide as quickly as possible whether the probe is a subfigure of the original shape. Generally speaking, the rate at which their response time increased with angle varied with practice, with the particular shape being rotated, and with the "goodness" (in the Gestalt psychology sense) of the probe figure relative to the rotated shape. Pylyshyn concludes that the holistic rotation of images cannot be considered an analogue process, since its rate depends on cognitive factors.

Let us dispel any suspicions of partisanship by pointing out that all parties agree that our account of piecemeal rotation remains unscathed by Pylyshyn's experiments, however they are interpreted. Nevertheless, we wish to take issue with some of the conclusions Pylyshyn draws from his findings. First, the finding that the increase of response time with angle of rotation decreases with the "goodness" of the probe figure does not necessarily imply that images were rotated at different rates when they were to be matched against different probes. It could be that the less "good" a subfigure is, the more that misorientation will complicate the task of "finding" that subfigure in the target. Thus greater slopes of reaction time versus angle for "bad" subfigures could reflect differences in the *comparison* phase, and not in the *rotation* phase. (This counterinterpretation could be ruled out by control experiments, of course.) Second, we also take issue with Pylyshyn's inclusion of practice as a "cognitive" factor supposedly penetrating the rotation process. Surely there are few cognitive processes at *any* level of description that cannot be sped up somewhat from their initial unpracticed rate. The explanation for this phenomenon (and others such as decay, fatigue, noisiness, and capacity limitations) may indeed refer to some property of neural tissue (in this case, perhaps the strengthening of synaptic connections with repeated firing), but this property is likely to cut across many or most primitive cognitive processes, since all such processes are probably instantiated in the same type of neural tissue. That mental rotation is affected by practice does not make rotation more cognitively penetrable than other processes. Finally, it is unclear that rate changes of *any* origin are examples of cognitive penetration. The rotation operation may indeed be a primitive process, but one of the arguments in its "hardwired instruction format" may be a rate parameter. People may choose in advance slower rates for "worse" probes, perhaps because a serial or capacity-limited process will be monitoring the rotating pattern to prevent it from fading or scrambling, or to decide when the rotation is to be halted. They then insert this rate parameter into an appropriate "slot" in the rotation instruction, and rotation ensues. Or, an additional stage might precede rotation proper, which sends pulses that initiate one increment of rotation. The rate of these pulses could vary depending on cognitive factors. The important point here is that the determination of the optimal rate is no doubt a penetrable process, but the *rotation* itself need not be.

Our second major point of disagreement concerns Pylyshyn's assumption that a cognitive process is not truly explained unless that explanation refers to primitive information processes. The argument against this position is nothing more than a special case of the argument against reductionism in psychological explanation (Fodor 1968; Putnam 1973). Though one can logically reduce a set of phenomena at one level of description to events at a lower level, one does not thereby *explain* those phenomena in terms of the events at that lower level. In Putnam's terms, we must take care not to confuse the "parents of an explanation" with the explanation proper. That is, let us suppose that all the functions and structures included in our account of imagery are not primitives but, like subroutines, or instructions in a high-level programming language, are compositions of more primitive processes. One subroutine would define the surface matrix (perhaps using the equivalent of a Fortran DIMENSION statement), another would fill its cells according to stored files of coordinates, and so on. Though we would not want to claim that this is the case in general, it is not wholly unreasonable to suppose that many cognitive processes might decompose into subroutines, the

subroutines at each level of decomposition selected mostly from a separate library, with the primitive operations comprising the lowest level library. As we pointed out in the section on Anderson's argument, a perspicuous account of mental rotation will refer to properties of a spatial structure, especially to the fact that orientation, size, and shape "reside" in the same representation in that structure. Whether the structure is itself primitive or is defined by subprocesses at a more primitive level, *the level at which the structure is considered spatial is the level at which mental rotation is explained.*

Finally, in closing this section it may be useful to consider which components of the imagery system are likely to be susceptible to cognitive penetration. It is a central tenet of most imagery researchers that imagery resembles perception (see Shepard and Podgorny 1978). Thus, one would expect that if a particular perceptual structure is cognitively impenetrable (as many must be, otherwise we would all literally be solipsists) and if that structure is among those shared by imagery, that aspect of imagery is predicted to be impenetrable as well. Not only will this allow us to provide accounts of the observed similarities between perception and imagery, but it constrains future theorizing in both domains, thus bringing closer the general goal of strong, explanatory theories.

Which components of imagery are likely to be common to the perceptual system? We will consider in turn the five classes of structures and processes that we have posited: the visual buffer, the long-term memory files, the image construction processes, the image inspection ("mind's eye") processes, and the image transformation processes. First, we suggest that the visual buffer is shared by vision and is probably the cortical medium underlying normal visual experience (see Finke and Schmidt 1977; Finke and Kosslyn, in press; Kosslyn 1978; and Pennington and Kosslyn, in preparation, for supporting evidence). Furthermore, we predict that this component will not allow cognitive penetration, that a person's knowledge, beliefs, intentions, and so on will not alter the spatial structure that we believe the visual buffer has. Thus we predict that a person cannot at will make his visual buffer four-dimensional, or non-Euclidean, or give it an arbitrary acuity profile, either in perception or in imagery. Of course, a person could conceivably *simulate* such properties in imagery by filling the visual buffer with patterns of a certain sort, in the same way that projections of non-Euclidean surfaces can be depicted on two-dimensional Euclidean paper. But this is different from changing the properties of the medium itself, which is what we doubt is possible. (Distinguishing among these possibilities experimentally will not be easy, needless to say.)

Two other imagery components may also do double duty in perception. Pure parsimony considerations lead us to believe that the processes that detect patterns in images (i.e., the "mind's eye" functions) will be the same as those that detect patterns in a spatial representation at some level of the visual system. Of course, little is known about visual pattern recognition, but often the distinction is drawn between "data-driven," bottom-up procedures, and "conceptually driven," top-down procedures. The latter procedures are those considered most likely to be cognitively penetrable, and we predict that if the distinction is valid, it will carry over to the "mind's eye" processes. Similarly, it would seem wasteful if the long-term image files are completely distinct from the stored representations used by the visual pattern recognition process. However, we can make no particular predictions at this point about whether or not these files are penetrable.

Finally, there are two sets of processes that we feel (1) are *not* shared with perception, and (2) *are* likely at least in part to be cognitively penetrable. These are the image construction and image transformation processes. As to the first assertion, people do not have to use special processes to construct or transform perceptual "images": the physical environment and their peripheral visual systems do that for them. As to the second, if imagery is to be a *useful* mental faculty, one that serves the higher reasoning processes, there should be some mechanism whereby what one believes, intends, hypothesizes, and so forth can affect the contents of images. After all, surely one of the functions of imagery is the ability to visualize the outcome of some operation that cannot, for whatever reason, be performed in the physical world (see Shepard 1978) – an ability that depends on interfacing imagery with general world knowledge. Thus we predict that whenever someone demonstrates a case of cognitive penetration of images, *some* – though not necessarily all – components of the construction and transformation procedures will be the processes penetrated. Kosslyn, Reiser, and Farah (submitted) have in fact demonstrated that one's motivation to construct a detailed image will affect in interesting ways the time taken to imagine an object at different relative sizes. Further, the locus of many of Pylyshyn's informal examples of penetration – such as the effects of prior knowledge on one's ability to perform color mixtures in images – almost certainly resides in the image construction or transformation processes. The reader should note, however, that the penetrability of the transformation and construction processes does *not* preclude the possibility that they are partly "analogue" in Pylyshyn's sense. For example, the transformation procedures might contain a primitive, analogue operation that translates or rotates a spatially contiguous region of points by a fixed amount. If so, cognitive penetration of image transformations could rearrange the sequence in which such operations are carried out, but could not redefine what each operation can do. (For example, in terms of our model, it would be impossible in a single step to select an arbitrary collection of dots scattered over the image and displace each dot by a different amount in a different direction.) In addition, even if these processes allow a degree of cognitive penetration, they may still form a structure of more-or-less cohesive subroutines, and, as argued earlier, the most adequate explanations of various cognitive phenomena may make reference to them.

In sum, we do not believe that Pylyshyn has weakened the arguments that imagery is a distinct "analogue" form of mental representation. Nevertheless, we accept his challenge to determine which aspects of the imagery system are computationally primitive and which are not, and to specify on which features of the theory the explanatory burden falls in accounting for various imagery phenomena.

3.4 Evaluating theories and models

In closing, we would like to reflect on how the present project should be evaluated. On our view, it is sometimes evaluated from the wrong perspective, because of a confusion between the proper respective roles of theories and models. The approach we have taken reflects a distinct view of theory construction. In a sense the task is well specified: devise an abstract framework that explains existing data and projects to new data. This task would be difficult even if data could be taken at face value. But what makes the task exceptionally difficult (and the preceding description of it so misleading) is that data cannot be taken at face value. Data can be problematic in at least two related ways. First, the proper extent of the data domain is not automatically clear. For example, should an imagery theory account for the ability of creative people to "see" solutions to problems in images? Should it explain eidetic imagery [see Haber: "Twenty years of haunting eidetic imagery," *BBS*, this issue]? Dreams and hallucinations? A *proper* domain is one that lends itself to theory construction. Second, it is often unclear which data should be discounted on the grounds that some *ceteris paribus* condition is violated. For example, could we reject embarrassing data on the grounds that *Ss* were highly practiced, or kept their eyes open when they formed images? The problem is that while factors external to the domain of interest can confound data at the outset of theory construction, the very distinction between "external" and "internal" factors is at issue. Of course, once a reasonable fragment of a theory has been developed, data can be interpreted and evaluated in a systematic way. The fragment of the theory can then be gradually refined and extended in the manner familiar in the mature sciences. But in a science like cognitive psychology, such a fragment of a theory is not available, and theory construction therefore requires some other means for interpreting data in a principled way.

Our view, then, is that there is a special problem at the early stages of theory construction. Various approaches to this problem are possible. For example, attention can be confined to an extremely narrow domain, within which data are quite well behaved – perhaps in our case, the rotation of imagined letters. The trouble with this approach is that it is not clear how to develop a general theory covering all the tasks within an entire domain, like imagery, by studying such narrow cases. Our approach to the problem was to use a partly metaphorical, sketchy protomodel to do the work that a fragment of a theory would do at a more advanced stage of research. The protomodel was then developed into and replaced by a more detailed process model. As with any other analogy, there were aspects of the protomodel that were to be taken literally and others that were clearly metaphorical (no one thought that humans have a cathode ray tube in their heads). Similarly, some aspects of the detailed model are to be taken seriously, and some definitely are not. For some aspects, however, it was initially unclear whether to count them as part of the "positive" (theory-relevant) or the "negative" (theory-irrelevant) analogy. Furthermore, the detailed model was not intended to be complete from the start, just as the initial protomodel was left open to further specification and elaboration. The research program, then, is now one of fleshing out the model in detail, in the process clarifying the ways in which it is mere metaphor and those in which it expresses substantive theoretical claims.

Such an approach in the early stages of theory construction is commonplace in the natural sciences: an excellent example is Maxwell's development of electromagnetic theory, using models based on the elastic deformation of continuous media. Although fruitfulness in research is an obvious justification for retaining a particular model, it has consistently failed to suffice for defending the model against complaints that it is being tailored to fit the data. Of course, this impression is not entirely inappropriate, for the model *is* consistently being revised and extended to fit the available data. In fact, not only do we agree that the model is often being made to fit available data, but we argue that this is the essence of the model development process. An important conclusion that should be drawn from this is that the fit between model and data is *not* the only consideration in evaluating a model. That is, the standard criterion for evaluating advanced scientific theories is successful fit with data, in particular with novel data or data of a significantly different kind from those used in devising the theory. We are cautious about placing great weight on this criterion in evaluating our model. Of course, we are pleased when the model yields predictions that experiments then corroborate. In fact, a *failure* of the model to make correct predictions would be disastrous. But we are reluctant to draw firm conclusions from the *successes* of the model, for the following simple reason: If the new data are in fact similar (in critical regards) to data initially used to motivate construction of the model, it is not very surprising that the model can account for these new data. At an advanced stage of theory development, novel data can be distinguished from old data in a different disguise, but this cannot be done with confidence at an early stage.

The criterion for evaluating our kind of model turns on the way the model is developed. Some experimental results will force it to be extended further by adding new data structures, processes, or properties thereof. Such extensions result in the model's having more *descriptive* power. Other findings will force it to be made more specific by contraining the nature of given data structures or processes, such as by fitting parameters (e.g. the amount of time necessary to retrieve and image the content of one "unit" stored in long-term memory). This will result in the model's having more *explanatory* power (e.g., once the parameter was fixed, the precise time required to form an image could be predicted, not just qualitative trends in the data). Thus the descriptive and explanatory power of the model are in tension, one trading off against the other.

The critical question to ask when the model is being extended or refined, is whether the change can be introduced without having to backtrack and modify parts of the theory that were fixed on the basis of earlier experimental results. And here, on our view, lies the appropriate metric for evaluating models: so long as research continues to yield a predominantly cumulative pattern of development – without the necessity to backtrack – one has reason to think that the model provides a basically adequate framework for interpreting data. But if experimental results consistently force arbitrary reformulation of the theory, it should be viewed with increasing suspicion. Our reply, then, to the charge that our model is ad hoc since we have to extend it in the face of new data, is that the wrong criterion – fit with data, particularly novel data – is being applied. This is the criterion by which relatively advanced theories are to be judged. Our model is being used primarily to make the theory construction process tractable in its early stages, and the appropriate criterion will reflect this role.

As we remarked before, models are ultimately not a good substitute for a theory. Models have gratuitous elements that remain unrecognized in the absence of a general theory. A simulation model is not the ultimate goal of our research program. At present, our simulation model is a tool in what is primarily an experimental research program. As the theory emerges, the detailed process model – that is, the simulation model – associated with it leads to interesting (i.e., precise or nonobvious) predictions, and hence further helps us collect data promoting the formulation and evaluation of the theory. Although the ultimate theory may not look anything like a computer simulation, we believe that the use of such a model will further the development of such a theory – if only by promoting systematic research in search of lawful regularities underlying empirical findings.

ACKNOWLEDGMENTS

Requests for reprints should be sent to S. M. Kosslyn, 1236 William James Hall, 33 Kirkland Street, Cambridge, MA 02138. The present work was supported by NSF Grant BNS-77-21782. The second author was supported by a NSERC Canada Postgraduate scholarship. We wish to thank Shimon Ullman for valuable suggestions and Debbie Lathrop and Sharon Fliegel for technical assistance. Steven Pinker is now at the Center for Cognitive Science, Massachusetts Institute of Technology, and Steven Shwarz is at the Department of Computer Science, Yale University.

NOTES

1. For the sake of readability, "he" will be used generically to refer to people of both genders.

2. This methodology can be contrasted with that of Shepard and Cooper, who often do not mention imagery in the task instructions. Instead, they define some visual classification task (such as deciding whether a misoriented letter is depicted normally or mirror reversed) that most *Ss* find is most easily solved by using imagery. *Ss* are told only to respond when they can classify the pattern correctly, and not when the image has met a certain requirement. The problem with Shepard and Cooper's technique (aside from its use of highly practiced and nonnaive subjects) is that it is not readily adaptable to those tasks that inherently admit of both imagery and nonimagery strategies, such as answering questions about properties of objects. Here imagery instructions must be used in some cases, so as not to combine responses that are produced by different underlying mechanisms, and so as to determine the exact differences between imaginal and nonimaginal thought.

Open Peer Commentary

Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article.

***by* Robert P. Abelson**

Department of Psychology, Yale University, New Haven, Conn. 06520

Imagining the purpose of imagery. The assertion that imagery is merely a special case of propositional knowledge sounds to me like the claim that Shakespeare's plays weren't written by Shakespeare, but by someone else who just called himself Shakespeare. While the imagery assertion could conceivably be true at some level of analysis, it draws attention away from the rich phenomena of imagery processing itself. Kosslyn and his coauthors have produced an impressively detailed model of how imagery might work, along with a spectacular array of supportive experiments. I find this much more satisfying than

the metaphysical sparring that has received so much attention in the literature, and I look for the whole imagery controversy to deescalate to the point where the issues are productive rather than distracting. It is a hopeful sign, for example, that Kosslyn et al. have acknowledged the cogency of Pylyshyn's penetrability criterion [see Pylyshyn: "Computation and Cognition" *BBS* 3 (1) 1980] and seek to articulate where it applies within their model, rather than dismissing it out of hand.

I found helpful the distinction the authors make between a "theory" and a "model." One of the greatest causes of confusion and skepticism about computer simulation models is that they generally contain so much detail, much of it seemingly gratuitous, that they apparently fly in the face of the principle of parsimony for scientific theories. As theories, computer simulations are complex and "busy." This may not in fact always be a bad thing. In my review of computer simulation in psychology (Abelson 1968), I defended the possibility that Occam's Razor might need modernization as Occam's Lawnmower. Still, the lack of parsimony in models disturbs many people. But now Kosslyn et al. tell us that one could have a theory much simpler than the model that exercises it, because features of the model recognized to be gratuitous or provisional should not be scored against the simplicity of the theory.

A similar point is often made informally by artificial intelligence workers in natural language processing by computer. They will say, for example, that psychologists should not pay attention to the low-level process details of whether stored knowledge is accessed in their models by serial search or discrimination nets, for example, because that's not as important as how the program knows what knowledge is relevant to look for in the first place. This kind of argument is well taken, provided the features of the model that are not to be taken seriously can be specified exactly.

This is not always easy. Indeed, the present authors do not exemplify their own model-versus-theory distinction very well in their model and theory of mental imagery. They don't tell us explicitly which details of their own model are too grimy to sully their theory. The reason for this, I think, is that the present version of the theory is too close to the model, arising as it did as a creature of the protomodel. As brilliant as this contribution is to the understanding of mental imagery in its own right, one would still like to see a broadening of the theoretical basis so that imagery is related to other aspects of organismic functioning. Such theory as is given here does not follow from general psychological functioning, but seems rather to be tailored to explain fine details of imagery processing itself.

The authors mention the question of how imagery relates to perceptual recognition, and that is indeed a fundamental problem to pursue. I would like to suggest two or three other problems at the boundary of the imagery domain that might cast a more functional light on just what this funny set of special processes (SCAN, ZOOM, FIND, ROTATE, etc.) might be doing in the human cognitorium. Tests of individual differences in the ability to form mental images show that some 10 to 12% of individuals claim little or no experience of imagery. (It seems that this experience may correlate highly with perspective on the controversy over imagery: Kosslyn tells me that he has very vivid images, and at least one of his theoretical opponents states that he hardly has them at all.) These subjects typically get averaged into the data in imagery experiments and introduce a little bit of noise. Or perhaps they are weeded out. But in either case, no account is given of how they function differently in tasks of the kind discussed in the paper. What are they missing if they don't have imagery? Can they not find their ways in rarely visited cities, nor assemble mechanical parts, nor play billiards well, nor have interesting dreams? What is imagery *for*, in other words, beyond deciding whether the horse's tail or knees are higher off the ground? Do nonimagers have some way to compensate in these activities by using nonimaginal processses? Of what sort? Or do they really image after all, but below the level of consciousness?

Another interesting group to examine closely in Kosslyn-type tasks would be blind people. If the parametric relations were generally the same as for sighted imagers, this would argue for a more abstract spatial representation than something based strictly on commonality with visual perception. (But the Kosslyn et al. model is pretty abstract anyway as embodied, for example, in their matrix for depicting a skeletal image. I am skeptical that spatial imagery is *necessarily* tied to visual perception. Can people image perceptual illusions?)

One final functional question which has for a long time intrigued me is, why do people typically report experiencing visual imagery when listening to stories? This very, very rarely helps at all in understanding what is going on (or so, at least, is the assumption in the story-understanding computer programs discussed by Schank and Abelson 1977). What, then, is its purpose? What are the general purposes of mental imagery? I hope that imagery researchers with the incisiveness and perspicacity of Kosslyn, Pinker, Smith, and Shwartz will some day tell us.

***by* John S. Antrobus**

Center for Research in Cognition and Affect and The City College of the City University of New York, New York, N.Y. 10031

Matters of definition in the demystification of mental imagery. I am impressed by the systematic program of theory construction, model building, and experimental research described in this paper. But I am less persuaded than the authors that any decisive positions have been established on such major issues as the analog-propositional controversy or on the epiphenomenon debate. I am reminded of the old argument as to whether an avalanche creates sound in the absence of a listener as I read that images "depict" information in a spatial medium, as opposed to "describing information" (sect. 1.1). I am suggesting that if we agree to the definition of the fundamental elements of visual and other kinds of cognitive representation, and to the definition of analog and propositional as they refer to those fundamental elements, the analog-propositional controversy regarding visual imagery will become trivial. The reader is encouraged to read Palmer's (1978) lucid paper on the subject.

The authors state (sect. 2.3.1) that "there are two sorts of representations used in the generation of images, one storing information about the 'literal' appearance of an object, and another... facts . . . interpreted semantically." This is a common method of classifying simulation models, but it sets up a straw man as far as cognitive theory is concerned. Any relational statement is a proposition. Even an x,y dot matrix, which is the primitive set of elements in the modeled image, is described propositionally in the computer program. (The CRT display is strictly epiphenomenal.) Further, even the simplest spatial shapes and forms such as a vertical line, angle, or curve must be described by relational statements, though not necessarily semantic propositions. Of course, semantic propositions must enter the model to interpret the experimenter's instructions, "Image an elephant," and to locate the image characteristics listed with the semantic term.

But these are criticisms of the controversy, not of the theory or the model. At this point, I say: forget the analog-propositional dispute and get on with the nitty-gritty evaluation of the dot matrix model. Compare the resolution and decay values under different conditions: after-images, visual-motor tasks, and spontaneous visual images as well as the procedures already employed. Compare the dot matrix with one using lines and curves as the primitive visual elements (Hubel and Wiesel 1962), and don't be distracted by disputes unless they can be defined in terms of the elements in your model.

With one exception. The epiphenomenon issue is important in that to answer the criticism one must show that some characteristic of the dot matrix, in this case, permits the simulator to extract information that was not available to the processor as long as that information was in another form, such as long-term memory (LTM). For example, spatial information about parts of an object may be available in LTM together with some information about relations to some other parts, but the formation of the image permits the computation of new information about the relations among the parts. The epiphenomenon criticism says that the image adds nothing to what can be computed by alternative processes. The evidence in the present paper is simply that there is information in the image, but no alternative models are compared.

Finally, I would like to suggest that the explanatory power of the proposed theory will be limited to the extent that it is a theory of visual imagery rather than of information processing. No living organism is

restricted to information in one sensory modality. Setting this artificial boundary will inevitably limit the kinds of behaviors that the authors attempt to model rather than increasing the power of the model.

by **Bruce Bridgeman**

Department of Psychology and Psychobiology, University of California, Santa Cruz, Calif. 95064

Neurologizing mental imagery: the physiological optics of the mind's eye. A brief look at the history of brain modeling in the context of information theory will help to put Kosslyn et al.'s model of mental imagery in perspective. In the 1930s, Turing showed that a very simple machine could compute any computable function: the Turing machine consisted of a tape and a mechanism that could write and erase symbols on the tape, read the symbols, and advance the tape, depending on the nature of the symbols. In 1943, in their landmark paper, "A Logical Calculus of the Ideas Immanent in Nervous Activity," McCulloch and Pitts showed that a Turing machine could be constructed of simplified neurons. A corollary of this principle is that there exist many configurations that can compute the same functions, so that an infinite number of models exists that can successfully simulate any consistent set of results.

After McCulloch and Pitts's breakthrough it was only a game to construct simulated nerve nets to perform any desired function, for it was known that a solution was possible, and it was only a matter of ingenuity to discover one. Since then the motivation of neurologically based modeling of brain processes, including cognitive processes, has changed to modeling a given function within the constraints of known neuroanatomy. Because a model constructed without regard to neuroanatomy is trivial, and a model that contradicts known principles of brain organization is not useful for understanding how brains work, it is necessary that models of cognitive function be consistent with the structure of brains and that they be homeomorphic: that is, that each component of the model should correspond to an identifiable component or property of the brain.

The use of physiological princples in cognitive psychology is not new: in fact, physiological psychologists have always been cognitive, for the last S-R oriented neuropsychologist was Pavlov. His S-R orientation was demolished by Lashley in the late 1920s, though the S-R orientation of nonphysiological behaviorists lingered in experimental psychology for another thirty years, largely because experimental psychology was paying too little attention to the physiology on which its explanations were ultimately based. This commentary is a plea that the same mistake not be made again by modern cognitive psychology.

When *Neisser* [q.v.] launched cognitive psychology in its present sense in 1967, he brought internal states out of the behaviorist closet and demonstrated some methods for examining them experimentally. Since then other powerful behavioral methods have supplemented his techniques. But psychology cannot lose sight of the fact that it is a branch of biology, the study of living things, and as such cannot become independent of anatomical and physiological knowledge. Thus the ultimate role of models must be the explication of the functions of brain structures, and modeling can no longer proceed without regard to brain structure. In fairness to Kosslyn et al., this may be asking too much of a young model, and their model has been useful in making their assumptions explicit. But the task of basing their model in neurophysiology and building homeomorphic constraints into it should not be delayed much longer.

Already cognitive psychologists are raising some issues that have no meaning when applied to the structure of the brains that in turn support the cognitive processing. An example from Kosslyn et al.'s article in Pylyshyn's concept of "cognitive penetration." This principle states that there should be "primary" brain processes that are not accessible to cognitive variables, and that are qualitatively different from higher cognitive processing [see also Pylyshyn: "Computation and Cognition" *BBS* 3 (1) 1980]. Real brains, however, show no obvious differentiation between primary and cognitive processing areas. Higher levels affect lower or more peripheral levels at every stage of processing. The physiological substrate for this process can be seen, for instance, in efferents from other brain areas to primary visual cortex (Spinelli and Pribram 1967), from there to the lateral geniculate nucleus of the thalamus (Singer 1979), and even into the retina itself (Spinelli, Pribram, and Weingarten 1965, Spinelli and Weingarten 1966). Similar efferent paths occur in other sensory systems. Thus no level escapes efferent control, and "cognitive penetration" becomes a matter of quality and degree. In general, there is no such thing as a one-way connection in the central nervous system, so that hierarchies are better termed "heterarchies," and anatomical levels of processing do not correspond to a concept of logically sequential stages.

What would a model of higher cognitive processes, such as that of Kosslyn et al., look like if it were constructed within the limitations of homeomorphism and anatomical consistency? The construction of such a model is far beyond the scope of this commentary, though some of the broad outlines of the model can be described. The model might be based on a "universal neuroanatomy," a conception of general brain structure that consists of a series of layers or sheets of interacting cells, each cell having parallel input and output connections to other sheets of cells above and below it. The brain would then consist of a set of layers made of receptor cells, a stack of parallel sheets of neurons with axonal connections between them, and at the other end a set of parallel outputs from the final sheet to the muscles. If one allows for some convergence and divergence, and if axons can occasionally skip a layer or two, most neuroanatomy can be described in terms of the simplified scheme.

To construct an anatomical model of experiential events, several more anatomical facts must be considered. First, most of the fibers entering and leaving the cortex originate not from peripheral or subcortical centers, but from other cortical areas along U-shaped fibers. Second, the most important cortical connectivity, at least on a gross level, is vertical in the cortex rather than horizontal. The evidence for this generalization is that cross-hatching the cortex with a knife seems to have little effect, while undercutting it, leaving the gray matter undisturbed, has the same effect as removing the cortex. The third consideration is the time discrepancy between the speed of individual neural events and the speed of reaction times and processing [see Wasserman & Kong: "Absolute Timing of Mental Activities" *BBS* 2(2) 1979]. Transferring a signal from one side of a synaptic cleft to the other requires about half a millisecond, and a reasonably fast axon can transmit an action potential from one side of the brain to the other in less than 10 msecs, yet reaction times are on the order of hundreds of milliseconds, and manipulations of mental images are no faster. This implies that the processing required for these tasks consists of many cycles of iterative activity, in which the state of a particular patch of cortex is broadcast out on U-shaped fibers to another piece of cortex, where it interacts in the graded-potential mode with the state encountered in that patch of cortex. This new activity in turn projects along U-shaped fibers to another area, or perhaps back to the same area again, in the action potential mode, and upon encountering its target cortex is convolved with the information present there. Thus neural information processing is conceived as a series of transformations of activity coded in combinations of many thousands of parallel neurons, undergoing incremental transformations as it moves from one cortical area to another. The incremental nature of Kosslyn et al.'s transformations of mental images might correspond to this sort of serial processing network. Evidence recently collected in my laboratory from single cells in the visual cortex suggests that the same cells can reprocess information in an iterative manner at least three times between stimulus and response, aside from the contributions of other areas.

The details of a model of imagery based on homeomorphic principles remain to be worked out, though a single neural imaging layer has been simulated (Bridgeman 1971, 1978). The broad outline given here is only a vague start on such an effort. But the combination of our exploding knowledge in cognitive psychology and new advances in neuroanatomy shows a promise of making possible much more explicit models of cognitive processing in real brains.

by **Lynn A. Cooper**

Department of Psychology, Cornell University, Ithaca, N.Y. 14853

Modeling the mind's eye. The research program of Kosslyn and his collaborators has clearly advanced our understanding of the nature of

mental images. While any single experiment in the series is vulnerable to criticism, the convergence of evidence over many experiments is impressive indeed. The empirical work reviewed by Kosslyn et al. has established that images generated by subjects upon request have properties like size and a characteristic spatial extent, and that such images can be subjected to operations like mental scanning.

At the core of Kosslyn's current effort to understand imagery is the computer-simulation model. The simulation model described by Kosslyn et al. does more than provide a convenient account of the data from the group's imagery experiments. In addition, the model is a serious first attempt at a theory of how information is represented in images and the nature of the processes that are applied to images. Kosslyn et al. offer some more or less standard justifications for embodying their model of imagery in a computer simulation. Among the most important functions of the simulation are that it ensures precision and that it leads to novel predictions deriving from complex interactions among components of the model that might not be obvious in the absence of the running computer program.

Despite these clearly positive features of simulation models, I have some worries about the long-run impact of the model on Kosslyn's empirical and theoretical work. First, it is not clear that the model has actually been necessary in arriving at some of the interesting distinctions made by Kosslyn et al. For example, the distinction between "field-general" and "region-bounded" transformations – while conveniently incorporated in the model – could certainly have been made in the absence of the simulation. Second, and more important, the new questions and predictions that the model leads to seem highly specific to this simulation and to the particulars of Kosslyn's experiments. Examples of these new questions include the nature and order of generation of parts of mental images, the degree of resolution of the visual buffer, and the extent of an image that can be "seen" at once without scanning the image. What I fear is that the next several years of Kosslyn's research program will be directed toward designing clever experiments to answer model-derived questions such as these while possibly ignoring other significant issues in the study of mental imagery.

Issues about which the simulation model is silent, but which to me seem important, might include: what is the role of imagery in a more general cognitive system? Under what conditions are mental images *naturally* used, and what is their function? What is the nature of the differences among mental images generated by different individuals or for different purposes? For example, are the internal representations and processes used to solve spatial problems or to generate a "cognitive map" of the environment qualitatively similar to those used to generate an image of a specific object for purposes of verifying whether or not it contains a particular feature?

In all fairness, Kosslyn et al. explicitly and carefully address the issue of the proper domain of their model and eventual theory. The only question that I have about their analysis is at what point the direction of research and theory will move from finely tuning the details of the model toward expanding the range of phenomena under consideration. Finally, these comments should not be construed as reflecting a general negative attitude toward simulation models. The advantages of such models, in particular the precision that they ensure, are obvious. And the particular model described by Kosslyn et al. achieves the goals that they set for it. My main reservation is that, as long as the simulation model is at the heart of their theory and research, the progress they make may be at the expense of ignoring additional challenging issues in the study of mental imagery.

by **Manuel de Vega**

Department of Psychology, Universidad de La Laguna, Tenerife, Spain

On interpretative processes in imagery. The traditional criticisms of imagery theories are that they are ambiguous and metaphorical. In this paper Kosslyn and his collaborators offer a computer simulation model of imagery, supported by considerable empirical data, which overcomes such criticism by its formal features. In this commentary I would like to develop some of my reflections concerning processes in the mind's eye.

First, I have some doubts about "image inspection" processes, because I feel they are indistinguishable from "construction" processes. I will start with an analysis of the latter. I do agree with the authors that images are not retrieved in toto but are elaborated or "constructed from parts," and, because of the structural limits of active memory, these portions fade nearly immediately, so that "effort is required to 'refresh' (them)." Thus, construction could be defined as a serial process whose temporal outline involves a set (N) of stages:

$$N = (S_1, S_2, S_3, \ldots S_n)$$

In the first temporal unit (S_1) only a portion of the whole image is depicted in active memory (perhaps a global or "skeletal" shape); in the next temporal unit (S_2) some details may be added, and so on. The construction process will stop at S_n when the image reaches the state required by the demands of a particular task. I dare say that the content of active memory at any given temporal unit is an unfinished product without functional value itself. Even at stage S_n the image could be lost because some previously generated details had faded. Accordingly, I suggest that the psychologically relevant or functional image – the complete product – is temporarily located not at any particular stage, but in the whole set N; that is, the image has a temporal extension.

When do the interpretative processes occur in this schema? One answer could be when the construction process has finished – immediately after S_n. If this were true, the efficiency of inspection processes could decrease drastically because at this stage, as at any other, the image depicted is a partial one. An alternative response is that inspection occurs all along the temporal set N as a process parallel to generation (this could prevent the loss of information and as a consequence the whole functional image might be interpreted). In this case, however, I find it superfluous to consider the "mind's eye" as a separate mechanism because the generative procedures themselves are semantic and interpretative in order to build a functional image adjusted to task demands. My claim is that interpretation (or inspection) is just a functional feature of construction processes. In fact, in their own model, Kosslyn et al. establish identical, or closely related, routines in both construction and inspection processes. For instance, FIND is a common procedure; and LOOKFOR is an inspection routine that activates the generative procedure IMAGE.

Second, I think that it is incorrect that "the processes that detect patterns in images (the mind's eye) will be the same as those that detect patterns in spatial representations at some level of the visual system." This speculative statement is forced by the imagery theorists' tendency to postulate functional and structural links between perception and imagery. But in this case the convergence, if any, is limited. I agree with the authors that interpretative processes in perception are poorly understood, but obviously they operate from the proximal stimulus generated in the early stages of processing (perhaps from unprocessed iconic memories). On the other hand the "mind's eye" interprets a high order product (not a retrieved iconic trace). The qualitative differences in informational material are enough to determine different interpretative processes between perception and imagery. In addition, pattern recognition is basically an input process, and the mind's eye is, in my opinion, a retrieval function.

In summary, I think that the conceptual distinction between generative and interpretative processes is justified neither by logical arguments nor by computation demands. Maintaining the duality, perhaps a theoretical bias of Kosslyn's protomodel, leads to a needless decrease in the parsimony of the model. Furthermore, I claim that the differences between interpretative processes in perception and imagery are substantive and result from specific informational characteristics of the two systems.

by **Jerome A. Feldman**

Computer Science Department, University of Rochester, Rochester, N.Y. 14627

So many models – So little time. The perspective of this commentary is that of an old-line robotnik turned recently to the behavioral and brain sciences and puzzled by the prevailing paradigms.

Much of Kosslyn et al.'s paper is concerned with the question of how one describes and examines theories and models of cognitive phenomena. Several years of extensive and beautiful experiments and

programming have culminated in some wistful statements about how little resolution of the imagery controversy has been achieved. It seems possible that extending the range of criteria for evaluating models could be of some help.

In addition to the usual scientific standards such as conformity with experimental data, parsimony, and generalizability, the major additional suggestion of Kosslyn et al. is that one should construct a running computational model. A convincing case for the value of these programs is presented, and many putative explanations are rejected because they were not stated with sufficient precision to be programmed. But we are all getting very good at writing programs and describing phenomena in these terms. The following two additional criteria have proved valuable in robot design and might be of some use to modelers.

Reducibility (Can the model be carried out by the substrate?). This does not mean that one must resort to neural modeling (although we may be ready for that). Analogous situations arise in astrophysics and cytology, for example. No one tries to build a detailed model of a galaxy or a cell from quantum mechanics, but theories that do not have a demonstrable reduction to physical reality have somewhat lower standing. Neither the TV metaphor nor the propositional model fares very well by this standard. For example, both employ the services of a homunculus, whose job I would have thought to have been automated some time ago.

Coherence and generality (Will it fit in with the rest of the system?). There seems to be widespread agreement that imagery and vision are related phenomena, and a good deal is known about vision at the physiological, psychophysical, and perceptual levels. The target article points out how imagery experiments constrain theories of vision, but does not use constraints from vision experiments in evaluating models and theories of imagery. The coherence/generality criterion suggests that the most basic knowledge about related activities be included in the choice of models. This does not require the construction of mega-models that attempt to explain all cognitive functions. The remarkable influence of such models must be due more to their overall structure than to their detailed scientific content, which undergoes continual change.

***by* Alastair Hannay**

Department of Philosophy, University of Trondheim, 7000 Trondheim, Norway

Images, memory, and perception. Like its main topic, Kosslyn et al.'s paper is a rather elusive one. The authors offer a *survey* of a theory of "image representation and processing." The survey is embedded in a more general discussion of the nature and proper beginnings of a theory of mental imagery, and this rests in turn on a set of assumptions about the form of an adequate psychological theory. The authors require of a satisfactory theory that it be able to "specify the 'functional capacities' of the brain" or "the various kinds of things the brain can *do* during the course of cognition." The latter formulation is important and revealing. No one suggests that mental images are themselves a form of cognition. The best one can do in the cognitive sphere is to incorporate imaging into a wider notion of imagination by giving it a role in the human ability to adopt a hypothetical frame of mind, to contemplate or envisage outcomes that it is impractical or undesirable actually to produce in the physical world, or in forms of 'mental' problem solving. The authors do not say whether specification of brain function is all that psychological theories should try to explain, or whether they mean that this at least is something such theories should include. But if mental capacities are brain functions, then imaging is indeed one of these, and a theory of the processes underlying mental imagery would be a legitimate topic for this kind of psychological theory even if imaging were as incidental to cognition as hiccuping is to eating.

The theory sketched is a sophisticated form of the traditional empiricist account of imaging as a postperceptual activity, an account that lends itself happily to the computer model that Kosslyn et al. use as the source of their theory. There are many points I would like to raise, but in this short space two will suffice. First, it is not clear to me that the terms of the theory, as they stand, do justice to the kinds of imaging the authors discuss. Second, it is not clear to me that the cases they discuss are the ones a theory of mental imagery concerned with cognitive functions and capacities of the brain should concentrate on. Let me say something very briefly on each point.

First, following Kosslyn et al.'s empirical findings of actual (not modeled) image performance, image "displays" do not give simple isomorphic renderings of stored data. They are products of processes that combine long-term memory units. In the simulation the images are generated on the matrix in steps, beginning with a skeletal image and then going into details. The model manages this with the processes named "picture," "put," and "find," and the authors hypothesize that brains have operations that accomplish the same ends. I think their case for this might be supported by referring to other mental accomplishments that are similarly concerned with placing parts into wholes. Locating a missing premise and drawing a correct conclusion are relevant nonvisual examples, and among the visual ones could be mentioned apt aesthetic placement of lines and colour patches in a drawing or painting, and their apt pictorial placement where the product is a (physical) representation. It seems very plausible that the same form of accomplishment can be found where the representations are mental. All one needs for the requisite theory is the spatial medium (or "visual buffer") for the image to appear in, and a capacity to store information about the look of things, along with other "propositional" information about visual aspects of the things with those looks. The authors (I think rightly) do not question the former postulate. Nevertheless, a full-fledged theory of mental imagery must allow for, even if it does not itself elaborate, a satisfactory account of the "visual buffer," particularly if, as here, the theory is designedly in opposition to those theories that claim that the spatial component is theoretically unimportant.

As for the existence in the brain of analogues of the visual memory files of the model, this, it seems to me, is not adequately explained by saying that mental images are products of the operation of processes on abstract representations (analogous to displays on a CRT produced by a computer program operating on stored data). This sounds far too much like the old "trace" theory of imaging to be true, even if the clothes are up-to-date. The paradigms for such a theory are visual wholes like chairs, cars, single letters or figures, or patterns or series of such, generated in an antihierarchical sequence with the abstract outline leading the way and the details following on. Of course this is a big advance on the old trace idea, for now the traces are sufficiently abstract to step into a wide range of pictorial roles, according to the processing possibilities (or rather capacities, since we are talking of what can actually be done). However, the more versatility we ascribe to the stored data, the less appropriate will it be to characterize imaging in terms of the retrieval of long-term memory units. The recombining and transforming will begin to bear the main explanatory weight, and the units that are processed into the image will resemble more the terms of a visual language than items stored in memory banks.

Finally, the authors say that their simple protomodel of the generation of images includes "interpretive" mechanisms that "work over" the internal displays and clarify them in terms of semantic categories. But if they mean this display to be the phenomenal image and not something even more internal (in which case "display" is hardly the right word), then a great deal of imaging will not conform to this model. Take the polar cases of deliberate visualizing (not the artificial cases used by the authors to map people's ability to form images and change them) and dreaming; in the former the imager is active hand in hand, as it were, with his or her brain, and in the latter the imager is a more or less helpless witness to whatever the brain turns up. These kinds of imaging, surely by far the most common, go no further than the "picture" process. They contain no phenomenal "putting" and "finding." There is, incidentally, an extensive and discriminating literature on the variety of imagings which psychologists might avail themselves of with benefit, even if it stems largely from armchair philosophizing (perhaps not an inappropriate location for fieldwork in this area; see Hannay 1971).

My second point concerns the terms of reference of a theory of mental imagery. Whatever the intrinsic interest of a theory of mental

representation, it seems to me that the more significant phenomena to be explained in the context of the cognitive capacities of the brain are those aspects of *perception* that, like representational imagery, are the brain's immediate contributions to the phenomenal content of the visual, or more widely, perceptual field. The cognitive implications of this can be suggested by reference to the notions of visual resemblance and experienced familiarity, both of considerable significance in terms of cognitive, and also adaptive, performance, and both highly complex qualities (or relations) which should come under the spirit, even if they do not come under the letter, of a theory of mental imagery.

by **Frederick Hayes-Roth**

The Rand Corporation, Santa Monica, Calif. 90406

Understanding mental imagery: interpretive metaphors versus explanatory models. Kosslyn and his colleagues have made two principal contributions to the psychology of imagery. First, they, along with others, have demonstrated that people can construct and manipulate memory representations manifesting realistic spatial properties. Second, they have developed a computer-based metaphor and a corresponding rudimentary implementation that simulates some hypothetical mental operations. While I fully endorse their methodologies and reported empirical results, the metaphor and model of mental imagery they have proposed seem to me both implausible and uninsightful. In this commentary, I briefly discuss several properties that seem to constrain potential models of imagery much more than the metaphor/model of Kosslyn et al. Subsequently, I propose some alternative models of imagery consistent with these observations. In the end, this contrast between past and future imagery models suggests desirable research directions and related theory-building approaches.

Imagery as perception: interpretation without explanation. The CRT (cathode ray tube)-based imagery metaphor rests on a central assumption that simultaneously makes the proposed model interesting and betrays its nearly total vacuity. This assumption states that mental image representations, like CRT-based displays, occupy a manifold that preserves spatial relations under homomorphism with physical space or some geometric projection of it. Said another way, mental images preserve relative distances between various parts of a scene. Moreover, image processing should exhibit a temporal dependence on interpoint distances that correlates directly with any corresponding distance effects in perception. In short, this assumption reveals an implicit model of imagery as visual perception modified to operate directly upon memory-generated, rather than sensor-generated, representations.

For this reason, the value of the proposed model depends on how well we understand perception, image generation, and image representation. None of these capabilities has been explained adequately or plausibly. As an example, consider the authors' supposition that image processing should proceed faster as the amount of information present increases: "Larger images are more quickly examined because more information is apparent." Although they argue as if this dependence derives from the spatial aspects of representation, it surely does not. Moreover the excessive generality of this presumption makes it easily falsified. In short, the appeal to weak notions of perceptual processes as explanatory conceptions will fail for several reasons. As in this example, so little is known about perception itself that few generally valid process characteristics have been identified. Many special characteristics of perceptual processing that are intuitively apparent will presumably occur in mental image processing too. Yet, correctly predicting such processing similarities does not explain or even "demystify" mental imagery. On the contrary, it merely ascribes to sensor-independent perception just those properties that we experience in everyday visual perception but which we do not in the least understand.

While I could consider at length several ways in which the metaphor/model of Kosslyn et al. offers ad hoc interpretations of phenomena in lieu of explanations, I mention just a few of these to motivate the subsequent consideration of more effective constraints on imagery theories. The preceding example exhibits what we might call an assumption that "large images speed processing." Later, in discussing mental rotation experiments, the authors argue just the opposite: "more time should be required to rotate subjectively larger images." In essence, "large images, slow processing." Presumably these opposing predictions do not derive from some common explanatory model, and commonplace examples could be found to support either proposition.

In another vein, the authors contend that because their model employs polar coordinates, it readily explains various empirically observed distinctions. Such an argument must presuppose, to explain any particular phenomenon, some further process-related assumptions about vector representations, multiplication capacities, or processor architectures. Thus, any explanation that rests on the assumption of polar representations does not actually derive from it directly. Furthermore, any such explanatory argument could readily be transformed into another coordinate system by merely replacing one coordinate set with another and one linear transform with another. I cannot see how the authors derive from general considerations or could support empirically one particular set of coordinates and related transforms. It appears then that they employ an arbitrary set in order to interpret (impute meaning or cause to) observed phenomena. In this regard, I view their interpretations as re-expressions of the data that employ unconstrained intermediate variables that could readily be replaced by a variety of equally interesting ones.

Major constraints on imagery and perception: representations and processes. Although their metaphor/model seems very weak, the authors have pioneered in a new and promising approach to imagery. The primary strength of their work, in my opinion, lies in the emphasis it places on depiction as opposed to description in encoding. This distinction continues a classic argument in psychology concerning the degree to which the mind encodes and manipulates images in analog forms. The authors specifically adopt a two-dimensional Euclidean basis for encoding images. In this space, interpoint distances are implicit in the spatial coordinates of the data elements. Because it captures spatial relations implicitly, the authors view such a representation scheme as depictive rather than simply descriptive.

Let us pursue briefly the nature of depiction to clarify the intrinsic properties that seem to make it theoretically interesting. I suggest that preserving spatial relations, as depictive representations are supposed to do, is obviously desirable for both computational and ecological reasons. Most generally, we should expect memory representations to preserve perceived relations that are required for movement and survival. Furthermore, we should expect the encodings of these relations to have evolved to expedite essential inference processes. This concept of depiction actually stands for the notion that data structures and processors have evolved together to store, access, and manipulate internal models of the environment.

The chief requirements of an organism's internal world model seem epistemologically obvious, and these should suggest natural constraints on encoding theories. Similarly, when coupled with common conceptions of human information processing, we can suggest plausible process models that would seem likely candidates to exploit such representations. I shall explore each of these issues, representation and processing, briefly in turn.

Representations, if evolved for efficiency and effectiveness in the organism's real-time self-control, should presumably encode a variety of environmental relations. In addition to distance relations, the representations of scenes should encode location, elevation, color, texture, object identity, size, orientation, adjacency, reflectance, and the like. As suggested by the authors, we would presume that representations should manifest these properties and relations inherently, rather than explicitly. That is, *the adopted representation should model the scene* to the maximum extent possible, rather than characterizing it in terms of a list of qualitative assertions. Thus, we conjecture that the primary constraint on hypothetical representations is that highly evolved representations maximize implicit information and minimize unnecessary derivative processing by modeling the object, scene, or phenomenon under consideration.

Image processing, on the other hand, would presumably have

evolved to enable rapid computation of significant inferences. For terrestial animals, as many researchers have pointed out, the primary needs are to recognize and locate targets, navigate toward them, and seize them (See Gibson 1966). By presupposing that the animal needs to keep its internal model current as it moves its eyes or body, we infer numerous constraints on the image processes. For example, they should enable real-time location and scale transformations that model movement of either the observer or the observed. To accomplish this, the transform processes must assimilate new coordinates for every depicted point in a scene. That is, the internal model must relate a constant set of objects and relations with a changing set of apparent visual loci. This militates for processes that establish correspondences between stimulus encodings and depictive models which can be maintained over a variety of field transformations. Thus, we should expect human image-processing systems to compute field transformations, over the entire image, in a fixed time unit. This implies both that such transforms would be computed in parallel over a distributed representation and that the intermediate states of an image under transformation should simulate an incremental physical alteration of the corresponding type. For example, in ways akin to previous rotation studies, we should expect orientation, scale, distance, shear, or stress transformations to be cognitively simulable and introspectively accessible.

Toward plausible models of imagery. As a model of human imagery, a computer-driven CRT display manifests few of these representation and processing characteristics. To shed light on either representation or process, we would need to extend such a model in every possible direction. We would have to specify hypothetical "software" that the computer would possess, the underlying "processor" that executes it, and the "recognition" procedures that detect and interpret objects in the eventually generated display. Traditional computer hardware and software provide a poor framework for pursuing these goals.

Recent advances in artificial intelligence, vision research, and computer graphics do suggest some promising ideas however. Work by Marr and his colleagues (Marr and Nishihara 1978) has focused on the development of depictive modeling representations called "2½-dimensional sketches" (corresponding to the "2½ dimensions" visible to an observer facing in one direction). Similarly, several machines have been designed and developed to compute field transformations (such as translation, rotation, scaling) very rapidly. Because these transformations effectively replace the contents of each cell by the contents of some cell whose coordinates are multiplicatively related to it, their speed is determined primarily by the number of multiplications that can be computed in a fixed time. Put another way, the principal transformations of imagery can be computed ideally by a memory that simultaneously associates all new and old memory locations under a multiplicative address relation. Such a capability can be built with today's hardware and presumably already exists in human brains for image manipulations (see also Marr and Poggio 1976, for a discussion of similar hypothetical architectures for stereo vision).

The primary characteristics of a proposed imagery model should include the following. (1) Representations should preserve immanent properties by exploiting homomorphic models; this means crucial spatial, temporal, and sensory properties should be modeled ("depicted") in the internal representation. (2) Real-time parallel field transformations should operate upon these models, as necessary, to update them to reflect observer or observed movement and other changes. (3) Perceptual actions, such as scanning, detection, recognition, classification, and hypothetical interpretation should be performed by operations that exploit naturally the distributed nature of the available representations and transformations (Hayes-Roth 1979). Moreover, these perceptual capabilities can be presumed to operate identically in purely mental imagery. This means, for example, that some visual comparison tasks could be computed efficiently by a series of successive transformations (e.g., translation, followed by rotation, scaling, and matching to compare two items). In addition, we should expect to discover a variety of capabilities that progress with practice from initially slow, sequential procedures of the sort just suggested toward procedures that eventually become instantaneous or act as real-time transformations. These transformations, which can be thought of mathematically as the composition of the initially successive transformations, would exploit and integrate two human memory capabilities – one for the parallel associative memory transfer that underlies the basic field transformations we have discussed, and a second one for compiling and foreshortening sequential procedures. In particular, these capabilities imply that practice will alter significantly the temporal properties of many task-specific imagery processes.

Conclusion: plausibility in imagery models. To develop plausible models of human imagery, one must begin with a proper appreciation of two facts. First, imagery derives from perception, and to understand imagery one surely needs to grasp the fundamentals of human perception. Second, human brains have evolved both to provide internal homologous mappings of external visual scenes and to process these scenes with elegant simplicity wherever high-bandwidth associative memory transforms suffice. In this regard, conventional computing hardware/software provides a poor metaphor either for visual representations or processes. An information-processing approach aiming to demystify mental imagery *is* warranted, but its success depends upon an adequate appreciation of the disparity between highly evolved natural machines and contemporary electronic ones.

***by* John Heil**

Department of Psychology, Cornell University, Ithaca, N.Y. 14853

Mental imagery and mystification. Suppose someone were to announce that he had discovered that mental images did not, after all, exist; that when people claimed to see things in their minds' eyes or hear tunes in their heads they were just wrong. Mental images, he might go on to claim, are on a par with ghosts and demons, holdovers from a metaphysically more tolerant past.

Such a "discovery" could not be taken seriously. It is true, if anything is true, that people see childhood scenes in minds' eyes, hear Sousa in their heads. We may wonder *what* mental images are – or, better, what imagining, the having of mental images, consists of – but not *whether* they are. Mental imagery is a fact one investigates, not a fact one seeks to establish.

In investigating and theorizing about a phenomenon, however, it is advisable to begin with a reasonably clear idea of what it is one is investigating. This is even more important when one's theorizing centers on the building of a "model" of the phenomenon in question. I shall argue that Kosslyn et al. are confused about the sort of thing mental imagery is and that this confusion infects their subsequent theorizing.

First, consider the two expressions:

i. *A* is looking at an *X*.
ii. *A* is imagining an *X*.

These expressions are superficially alike. This may be what leads us quite naturally to views that take imagining to be forms of interior looking or "scanning." I shall argue, however, that the mind's eye cannot be an internal analogue of an ordinary eye.

A crucial logical difference between looking and imagining is that *looking* is a relational expression and *imagining* is not. To say that *A* is looking at *X* is to say that *A* is in some relation to something, *X*. If it is true that *A* is looking at *X*, then it is true also that there is an *X* at which *A* is looking. This is not obviously so, however, for imagining. If *A* is imagining *X*, then *A* is not in some special relation of imagining to something, *X*. Nor need it be true that there is an *X* that *A* is imagining. Imagining is, as it were, logically intransitive; it is unlike looking and seeing: more like sitting and sleeping. To say that *A* imagines *X* is just to say that *A* is doing something, that *A* is imagining *X*-ly. The difference, then, between *A*'s imagining *X* and his imagining *Y* is not that there are two things, *X* and *Y*, to which *A* may be in some relation, but that in the one case *A* is doing something *X*-ly while in the other case he is doing something *Y*-ly.[1]

Admittedly, the idiom is somewhat awkward, but it is parallel to our way of speaking about painting, poetry, and fiction, not an unpromising parallel when one thinks about it. Consider a painting of a clown. If the painting is a portrait – that is, a representation of a particular clown on

a particular occasion – then to say that the painting is *of* a clown is to say that it stands in a certain relation to a particular clown. If the painting is not a portrait, however, then saying that it represents a clown is simply to say that it is a clown painting – roughly, that it is a member of a set of things collected together because they resemble one another in various ways (like tables and chairs). Such sets are characterized by similarities or family resemblances. Compare these to sets of objects defined by reference to their *denotata,* the set, for example, of drawings of Jimmy Carter. (see Goodman 1968.)

The nonrelational character of talk about the having of images accounts for features of images that otherwise might seem puzzling or mysterious. One cannot have an image of *X* without knowing that the image *is* of *X*.[2] Not so for looking, for example. You are asked to imagine a house by a pond in rural England. You do this. Would it make sense to ask whether you are certain the image is of a house and not, say, of a papier mache mock-up; whether you are certain the scene is in England and not in Arkansas? Imagine Jimmy Carter. Now: how do you know it is Carter and not his twin or someone else disguised as Carter? This shows, too, that images are not like pictures (where these and similar questions make sense). Having an image is like *drawing* a picture or a diagram, not like observing one.

Kosslyn et al. liken images to pictures. They point out that pictures do not represent what they represent just in themselves; they must be *interpreted.* A picture of a boxer in a certain pose might be used to show someone how to stand when boxing, or how *not* to stand; how someone stood on a particular occasion, or how one might stand. In recognizing this, the authors recognize that mental images cannot be modeled just on interior CRTs (cathode ray tubes), but only on these together with additional mechanisms whose job is to interpret the CRT display in some definite way.

Such a view is in danger of slipping into a bottomless regress. I shall not discuss that possibility here, however. It raises technical issues in the theory of representation that go beyond the scope of these remarks. It will be enough to point out that the theory is evidently at odds with our strongest intuitions about mental images.

Much of the point of postulating an interior CRT analogue rests on the failure to appreciate the nonrelational character of *imagine* and its cognates. The resultant feeling is that, if *A* imagines *X*, there must be some *X* that *A* imagines. Call this *X* a mental image and model it on a CRT. In this way a special class of entities is born, the metaphysical offspring of an unholy coupling of bad epistemology and I.B.M.

The fact that pictures must be given an interpretation – a fact recognized by Kosslyn et al. – raises serious doubts about any theory that models mental imagery on pictures of any sort, whether wax impressions (popular among Greek theorists) or CRT displays. Pictures, for example, but not mental images (or, for that matter, thoughts) may be ambiguous. The ways in which we can go wrong in misapplying pictures are not at all like the ways in which we can go wrong about mental images.

We may feel that the having of a mental image is like perceiving, but we may have difficulty in saying how. How, for example, is imagining an *X* in one's mind's eye like looking at an *X*? Not, I have suggested, in the sense that in both cases there is an object, *X*, that is observed (or, in a sexier idiom, "scanned"). Still, it may be that imagining an *X* involves knowing how *X* looks. Imagining one's grandmother is to imagine how one's grandmother looks (or looked, or might have looked).

This cannot be quite right, of course. One can imagine things that one has never seen and that, in consequence, one could not know the looks of. But even here, to imagine such a thing seems to involve a belief or guess about how the thing would look (or sound, or feel, or smell, or taste).

This ties imagining to the capacity to recognize things. We put this capacity to use in ordinary perception in the process of recognizing what we perceptually confront. In order to see *X*s I must, in some sense, know what *X*s are so that when I am visually confronted by an *X* I can recognize that this is what I am looking at. The ability to do this is, I want to suggest, at one with the ability to imagine *X*s, to have mental images of *X*s.

Have I, despite what was said at the outset, denied that there are mental images? Not at all. What I have argued is that (i) to have a mental image is not to be in some relation to some item having the properties attributed to the image; and (ii) the having of mental images is not like looking at CRTs. What I have denied, then, is that a particular conception of mental imagery is correct. I have tried to replace that conception with one that, I think, makes sense and is in harmony with our pretheoretical intuitions about mental imagery. These remarks, however, are logical, not psychological. The connection between imagination and recognition (i.e. perception) may be logical, but the character of our capacity to recognize is an empirical matter.

NOTES

1. Ulric Neisser's account of mental imagery (Neisser 1976, ch. 7) is an example of an empirical theory consistent with these logical features of *imagining.*

2. I leave aside cases in which one imagines things without knowing what they are called. The argument can, I think, be extended to cover these.

by **Geoffrey Hinton**

Program in Cognitive Science C-009, Center for Human Information Processing, University of California, San Diego, La Jolla, Calif. 92093

Imagery without arrays. Kosslyn et al. suggest that during visual imagery, stored knowledge is used to create an internal representation that is something like a 2-D array. They believe that this array could be used for several kinds of computation. It would allow the subject to notice new relationships that were not explicitly represented in the stored knowledge from which the array was constructed. It would also allow the subject to simulate changes in the real world by applying a sequence of small transformations to the contents of the array.

I shall argue that the array theory of imagery is wrong, and that the computations for which the array is invoked are more conveniently performed by manipulating nonlinguistic structural descriptions in which nodes represent objects or their parts, and labeled arcs represent 3-D spatial relationships.

The structural description theory. To see the physical world as stable and to recognize familiar objects from new viewpoints, it is necessary to have representations of spatial structures that do not change as the viewpoint changes. Such representations can be achieved by describing spatial structures relative to "intrinsic" frames of reference that are embedded in external objects. Changing the choice of intrinsic frame for an object may radically change its phenomenal shape. A square tilted at 45° is phenomenally different from an upright diamond. Attneave (1968) and Rock (1973) provide compelling evidence for the psychological reality of intrinsic frames of reference. Marr and Nishihara (1978) suggest that for complex objects (e.g. the human body), we use intrinsic frames of reference at several levels, so that the whole person has one intrinsic frame, whereas his arm may have a quite different one, whose relation to the frame for the whole is explicitly represented. Such structural descriptions are useful precisely because they are viewpoint-independent, but this makes them seem like poor candidates for visual images that are normally committed to a specific viewpoint. However, I shall argue that inferences are facilitated by attaching to each object node information that specifies how the intrinsic frame of reference of the object is related to the current, viewer-centered frame of reference. This "projective" information, which gives the structural description an implicit viewpoint, would be necessary to fill in a 2-D array from the structural description, but the array is redundant because it is the projective information itself that is needed for making inferences, not the 2-D array.

To represent the position and orientation of an object, it is necessary to represent the relation between the assigned, intrinsic frame of reference of the object and some other frame of reference. Two rather different kinds of relation are needed in vision. An "intrinsic" relation defines an intrinsic frame in terms of some other intrinsic frame. A "projective" relation, on the other hand, defines an intrinsic frame of reference by its relation to the viewer-centered one. Projective relations must be known to get from low-level, viewer-centered representations of the visual input to a structural description that uses intrinsic relations.

Some of the relations between frames can, theoretically, 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 relation between two intrinsic frames from information about the projective relations of the two frames to the same viewer-centered frame. This kind of computation must occur when we "just see" an intrinsic relation.

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 projective relation for the part must be computed from the projective relation for the whole and the known, intrinsic relation between the whole and the part.

A method of computing unperceived spatial relations. If a system stored knowledge of the spatial structure of a complex object and the object is not visible, then some computation must be performed to answer questions about relations that are implied by the stored knowledge, but not explicitly stored. This computation can be performed in two stages:

1. Choose a projective relation between the viewer-centered frame and a node in the structural description, and use the intrinsic relations between nodes to propagate compatible projective relations to the other nodes. On the structural description theory, this is what is involved in "working up" an image from a particular viewpoint.
2. Use the projective relations associated with the two nodes in question to compute the intrinsic relation between them.

This method enables a system that uses a hierarchy of intrinsic frames to compute an unperceived relation by making use of the mechanisms that must already be available for normal perception. The introspective evidence that visual imagery involves commitment to a particular viewpoint can now be interpreted as evidence that we chose a 3-D, viewer-centered frame in which to perform computations. This does not imply that we perform any of the hidden-line-removal computations that would be necessary for generating a 2-D array.

Transforming visual images. Kosslyn and Shwartz (1977) have tried to implement a system that performs transformations on the contents of 2-D arrays. They restricted themselves to dilation and translation, so they did not get bogged down by the difficulty of rotating an object about an arbitrary point in a rectangular array. Even so, their program demonstrates some significant difficulties for their theory. To explain why details disappear as images are shrunk and reappear when they are enlarged, two separate arrays are used, one of which is a blurred version of the other. One wonders why the homunculus doesn't peek at the clear image behind the scenes. Their only explanation for why transformations should be continuous is that points in the image are shifted sequentially, and so large shifts would cause a "noticeable gap." They would have done better to make each cell in the array a processor that can only access its neighbors, though this scheme has its own problems. Their computational model is inelegant and unconvincing for the easiest transformations, and it is hard to see how it could be extended to handle 3-D transformations such as rotation in depth, which people find just as easy as rotation in the picture plane (Shepard and Metzler 1971).

In the structural description theory, mental transformations involve changing some of the parameters of the intrinsic and projective relations, while leaving the rest of the description intact. The parameters that are changed are just those that would change during perception of a changing scene. A mental rotation, for example, is performed by changing both the intrinsic relation between the rotating object and its context, and the projective relation between the object's frame and the viewer-centered one. The parameters of the spatial relations are continuous variables, so the continuity of the transformations is easily modeled without appealing to spatially isomorphic representations, though it is not clear why it is easier to make continuous internal changes than discrete ones. The answer to this puzzling question may well lie in the way continuous variables and constraints between them are implemented in the brain.

The need to zoom in on fine detail, which Kosslyn et al. use to support the array theory, can be more economically explained in terms of structural descriptions containing continuous variables. It is reasonable to suppose that there are small errors in the representations of the values of variables. So if two variables have very similar values, the proportional error in the difference between them will be very large. If an object is imagined very small, the projective relations for its parts (their coordinates in the visual field) will only differ very slightly, and so it will be hard to compute new intrinsic relations accurately. It would be advisable first to increase the differences between the projective relation, that is, to increase the imagined visual angle of the object. The constant, known intrinsic relations could be used to maintain consistency between the projective relations as they are changed. The intrinsic relations are thus the source of the increased accuracy created by zooming in. Unlike the array theory, no extra representation is needed to explain the increased accuracy.

Evidence against the array theory. If a visual image is like an unsegmented 2-D array, then people should be able to reinterpret ambiguous visual images. In very simple cases this can be done. For example, an imagined capital N on its side can be reimagined as an upright Z. However, it is usually very hard to reparse complex, ambiguous images (Reed 1974, Hinton 1979). Imagine the outlines of two equilateral triangles of equal size, one upright and one inverted with its tip resting in the middle of the base of the upright one. How many parallelograms can you see in this configuration? People often find the central diamond, but the other two parallelograms are much harder to detect in the visual image than in a picture, because the picture can be reparsed much more easily.

The structure described above in terms of two equilateral triangles can be given an alternative structural description as two overlapping parallelograms that slant in opposite directions and have collinear ends. Using one structural description, it is relatively easy to visualize the triangles separating along a vertical axis, and using the other structural description it is easy to visualize the parallelograms separating horizontally. It is very hard, however, to visualize the same physical transformations when using the inappropriate structural description. Presumably, this is because simulating the physical transformations involves changing the spatial relations that are explicitly represented in the structural description, and many more changes need to be made to the intrinsic relations in the inappropriate structural description.

It is hard to explain these effects on the theory that mental transformations are generated at a level of representation below that of structural descriptions.

***by* Ian M. L. Hunter**

Department of Psychology, University of Keele, Staffordshire ST5 5BG, England

Mental visualization in nonlaboratory situations. Talk about the mind's eye and its ability to inspect mentally visualized scenes has long been common. Kosslyn et al.'s paper starts with the brilliantly simple idea that such talk is not always metaphorical and may sometimes be taken at face value or nearly so. The paper then reports a theory-building enterprise that has suggested psychological experiments that have yielded findings about properties of mental visualization: I might add that Beech (1979) has replicated some of these experiments and their main findings. In the paper, the findings are used to further the authors' theory building, and this is entirely proper. However, we may also consider the findings in their own right and ask, not only about their robustness, but also about their generality. Do they carry outside the laboratory to performances in which people report using visualization as a help in some goal-directed pursuit? Are there, in short, real-life situations for which the cited experiments are paradigms? I think there are, and I shall confine my comments to citing a few examples.

The method of loci (Yates 1966; Hunter 1977) instructs us to visualize a basic sequence of *loci* or places, to translate each presented item into a visualized image, and to knit each successive image to its corresponding place by combining the two in a visualized composite scene. The key claim is that mental visualization is a mode of representation in which associative learning is readily accomplished. Also, advice is often given about the spatial characteristics of effective

visualizations. For example, Willis (1621) advised that each place be a room that is open toward the viewer. The room is six yards wide and deep and high "and such a fashioned Repositorie are we to prefix before the eyes of our mind . . . supposing ourselves to be right against the midst thereof, and in the distance of two yards therefrom" (p.8). When placing an image in such a place, Willis advises adjusting the size of the image to ensure that it is "neither so great, but that it may be contained in one of the places; nor so small, but being there bestowed it may easily be seen by one that standeth two yards on this side of the [place]" (pp. 14–15).

In modern times Luria (1968) has given an account of a man, Shereshevskii, who used the method of loci, often in the form of a visualized mental walk. The man became a professional mnemonist, and we learn about ways in which, over the years, he modified his procedures to make them more efficient [see also Haber, this issue]. Two modifications relate to the paper by Kosslyn et al. One concerns the distance mentally travelled between one locus and the next.

"Earlier, if I were asked to remember the word *America*, I'd have had to stretch a long, long rope across the ocean, from Gorky Street to America, so as not to lose the way. This isn't necessary any more. . . . I'd set up an image of Uncle Sam. . . . I don't go through all those complicated operations any more, getting myself to different countries in order to remember words." (p.43).

The other modification concerns the complexity of visualization.

"Formerly, in order to remember a thing, I would have to summon up an image of the whole scene. Now all I have to do is to take some detail. . . . Say I'm given the word *horseman*. All it takes now is an image of a foot in a spur. . . . So my images have changed quite a bit. Earlier they were more clear-cut, more realistic. The ones I have now are not as well defined or as vivid as the earlier ones. . . . I try just to single out one detail I'll need in order to remember a word." (p.42)

A final naturalistic example concerns the use of visualization to serve temporary working memory. In a tape-recorded session, Hunter (1962) asked A. C. Aitken to recite from memory the value of Pi to a thousand places. Aitken did this without error, reciting at a rate of about five digits per second with a half-second pause between blocks of five digits. He then undertook to recite the digit sequence backwards, starting from the thousandth digit. Whereas his forward recital of the final fifty digits took a total of 18 sec, his backwards recital of the same fifty digits took 34 sec. He reported that, in backwards recital, he was "forced to use visualization." "I brought [the digits] along in blocks of five and managed to see them as a whole and read them backwards." He said that his forwards recital was an auditory-rhythmic activity involving no visualization: "seeing would put me off" and slow down recital. His self-reports checked with his observable recitations to suggest that, in the backwards task, he introduced visualization as a resource that enabled him briefly to arrest each successive five-digit block so as to perform the extra operation of reciting it backwards.

The point of my commentary is to suggest that the findings cited by Kosslyn et al. concern properties of mental visualization that are sometimes exploited by people in nonlaboratory tasks. If this suggestion is correct, laboratory explorations of visualizing will help us to interpret more securely the often puzzling characteristics of some real-life performances, and of the self-reports we receive from the performers. Reciprocally, naturalistic studies of real-life activities will help to establish the generalizability of findings from the laboratory.

by **P. N. Johnson-Laird**

Centre for Research on Perception and Cognition, Laboratory of Experimental Psychology, University of Sussex, Brighton, BN1 9QG, England

The "thoughtless imagery" controversy. Everyone agrees about the existence of mental images; the question at issue is the nature of their underlying representation. Kosslyn et al. argue that imagery depends on a special sort of representation, which they are at pains to distinguish from other sorts. Their opponents claim that there is a uniform "propositional" format for all mental representations, and that the subjective experience of imagery is a mere epiphenomenon. The disagreement should be a simple empirical matter, yet despite the numerous experiments addressed to it, it has yet to be resolved. Indeed, it has, like its illustrious predecessor, the debate about "imageless thoughts," polarized the participants into two camps that seem to be so entrenched that nothing will make them give way. In psychology, old controversies never die, they merely fade away . . . with the rise to fashion of other issues. What is unfortunate in the present case is that a greater insight into the psychological phenomena might have been gained if the dispute between the protagonists had never begun. Plainly, there are some nasty conceptual problems mixed up in it, and to try to disentangle them, I shall be forced to take issue with both Kosslyn et al. and their opponents.

There is a very simple and direct demonstration that only one mode of mental representation is necessary. The argument requires two assumptions: first, any "effective procedure" is computable by a Turing machine, and second, any adequate theory about a cognitive process constitutes an "effective procedure." It follows that any adequate psychological theory can be represented as a Turing machine. This device, of course, uses a uniform mode of representation – a linear string of symbols from a finite alphabet – and can itself be characterized in the same code by a set of propositions that specify its behavior as a function of its current state and the symbol that it is scanning. A Turing machine is thus preeminently a propositional device: it makes use of propositional representations and can itself be completely described by them. Moreover, granted the two assumptions above, it can carry out all the operations postulated by any psychological theory. Q.E.D.

Is this the point that the critics of imagery wish to establish? It seems unlikely that so trivial a matter is what really concerns them. Nevertheless, much of their argument has exactly the same form. When they suggest, for example, that a shape can be represented by a set of propositions about the locations of its various constituents, they are merely redescribing the shape at a lower level of description. They might as well extend this reduction down to the level of a Turing machine. In short, neither party to the dispute appears to have appreciated that the simplest case for propositional representations has no empirical content whatsoever, and that many of the arguments in the literature equivocate on this point. Until the notion of a propositional representation is given an empirical content, no experiment, however ingenious, will make any impression on a dedicated propositional theorist. The reader should not imagine that only the advocates of propositional representations are to blame. When Kosslyn et al. remark: "an image *must* be of *some* size and orientation," is their assertion to be interpreted as part of their definition of an image, as a logical truth, or as a testable hypothesis?

The "thoughtless imagery" controversy, as I have dubbed it, has concentrated attention on issues that have narrowed the scope of inquiry. The argument has focussed on images and propositions as exhaustive sorts of representation: there may well be other sorts including complex mental models that underlie cognitive performance without necessarily emerging into consciousness. If research had not been so concerned in demonstrating the existence, or nonexistence, of unique characteristics of imagery, then it might by now have established a better description of the range and variety of mental representations [see Pylyshyn: "Cognition and Computation" *BBS* 3(1) 1980]. An example of just such a constraining effect is to be found in Kosslyn et al.'s paper. The authors report that subjects who did not use imagery were faster to verify closely associated properties (e.g. that a cat has claws) than subjects who were given imagery instructions; but no attempt appears to have been made to follow up the nature of the mental representations used in place of imagery. Likewise, the emphasis on the putative distinguishing characteristics of images has led theorists to overlook the fact that modes of representation can in principle differ solely in terms of their function. Model-theoretic semanticists often use a set of sentences to serve as a *model* for a set of sentences; computer scientists know that certain lists can serve either as data to be operated on by some procedure or else as procedures themselves. The crucial distinction is in function, not in form or content [cf. Fodor: "Methodological Solipsism as a Research Strategy in Cognitive Psychology" *BBS* 3(1) 1980]. Hence, a propositional representation encoding a verbal description might be evaluated with respect to a propositional representation encoding an image. The two representations would nevertheless be distinguishable, but only in terms of their function.

High-level programming languages make use of a variety of repre-

sentations including lists and arrays, not because these increase the power of the system or necessarily allow it to run more efficiently, but because they make the programmer's task of developing and testing programs easier. If the mind can develop programs of its own (see Miller, Galanter, and Pribram 1960), then it too can use a variety of high-level representations for the same reason. Of course, ultimately they may all be translated into the "machine code" of the brain – just as one can reduce all cognitive theories to Turing machines – but such reductions should not be allowed to obscure the fact that very complicated programs can be developed only by working at a high level of description. One psychological example must suffice: there is evidence that the heuristics governing ordinary inference require a reasoner to construct a mental model that contains representative individuals. For example, to draw a conclusion from such premises as, "Some of the men in the room are monetarists, and all monetarists are followers of Milton Friedman," one imagines arbitrary numbers of individuals representing the relevant sets of men, monetarists, and followers of Friedman. Although it might be simpler to operate with rules of inference that apply to propositional representations of the premises, this method is evidently not used by logically naive subjects, presumably because they have been unable to develop the necessary rules [see Johnson-Laird 1979a).

Finally, let me touch upon one point that arises in Kosslyn et al.'s paper and that I have discussed in detail elsewhere (Johnson-Laird 1979b). Anderson (1976, 1978) has shown how a theory of mental processes making use of one sort of representation can be mimicked by a theory using a very different sort of representation, provided that the two theories classify stimuli into the same sets of equivalents. This demonstration has led many commentators to conclude that questions of representation are unlikely to be resolved. However, this conclusion should be treated warily. Here is an example of a theory that makes use of two sorts of representation that cannot mimic each other. The theory assumes that there are propositional representations couched in a mental language that is very close to natural language, even perhaps to the extent of a one-to-one mapping between their vocabularies (see Kintsch 1974; Fodor, Fodor and Garrett 1975). It also assumes that mental models that encode spatial relations in arrays can be constructed from such propositional representations. Hence, a description such as, "The cup is on the right of the saucer that is to the left of the plate," gives rise to a single propositional representation, but to several distinct mental models because the relation between the cup and the plate is indeterminate. In general, a mental model contains far more information than a propositional representation because the model can be constructed only by going beyond the information given in a verbal description: a picture may be worth a thousand words, but the meaning of ten words may be capturable only in an infinite set of alternative models. It follows that a finite set of models is consistent with fewer states of affairs than a propositional representation, and accordingly that the two sorts of representation do not yield the same equivalence classes for verbal descriptions. Once one has constructed a particular model, it is impossible to recover the original premises on which it is based. There is in fact a considerable functional advantage in using both sorts of representation: propositions encode indeterminate information economically; models facilitate the manipulations required by inferential heuristics.

***by* Janice M. Keenan and Richard K. Olson**

Department of Psychology, University of Denver, Denver, Colo. 80208 and Department of Psychology, University of Colorado, Boulder, Colo. 80302

The imagery debate: a controversy over terms and cognitive styles. One of the most striking aspects of the current debate over the status of mental imagery is that so much has been written with so little effect. As Kosslyn et al. note, neither the arguments nor the counterarguments has had enough force to sway people from their original stance on the issue. This is a frustrating state of affairs. Much worse, it sets the stage for radical proposals, such as Anderson's (1978) claim that issues of internal representation are fundamentally undecidable given only behavioral evidence. Although such a claim is tantamount to calling a moratorium on much of the research in cognitive psychology, it has nonetheless enjoyed fairly wide acceptance. We believe this acceptance is due to the fact that it provides justification for the failure to resolve the imagery issue: the controversy continues despite all this effort because the issue is fundamentally undecidable. But issues of representation are not undecidable. Several recent papers show that Anderson's argument to this effect is basically wrong (Keenan and Moore 1979; Pylyshyn 1979, in press).

If issues of internal representation are not undecidable, then why has there been so little progress in settling the controversy? Note, the lack of progress refers to the debate on the status of imagery as an explanatory construct in mental processing. Certainly, the results of Kosslyn's research program represent a tremendous advance in our knowledge about the behavioral concomitants and consequences of imagery. He has provided not only a chronometry of mental imagery, but also a broad empirical basis for the oft-voiced claim that the operations of imagination are highly similar to the operations of perception.

One reason the controversy continues is that the pro-imagery camp continues to use terms and analogies that are laden either with strange metaphysical connotations or with implications that contradict existing knowledge about psychological processing. For example, consider Kosslyn et al.'s definition of images: "Images are temporary spatial displays in active memory that are generated from more abstract representations in long-term memory. Interpretive mechanisms . . . work over . . . these internal displays and classify them in terms of semantic categories." There are several problems with this definition. First, by equating images with spatial displays, Kosslyn et al. imply that images are internal objects at which some homunculus can look. This raises the problem of an infinite regress of homunculi. Furthermore, it renders the image static, implying that parts of the display (image) exist whether or not they are attended to. But there is good reason to believe that, say, when one is imagining a walrus, the tusk must be constructed, not simply looked at, when one shifts attention to it (*Neisser* 1976, 1978). Second, by saying that interpretive mechanisms are needed to work over the display, Kosslyn et al. imply that the image or the abstract representation from which it is generated is basically sensory, that is, uninterpreted. This notion is reinforced by their building into the simulation a memory for the literal appearance of an object or scene, where the literal memory is not interpreted semantically. (To add to the confusion, Kosslyn et al. put the term "literal" in quotes, and never explain what they mean by it.) But, anything that is stored in memory has already been interpreted in some fashion; we simply cannot store raw sensory data. When we store appearances, they are stored as the appearances of objects or scenes.

Further terminological problems are apparent in statements like the following: "These results, then, support the claim that the images we experience are spatial entities and that their spatial characteristics have real consequences." To say that images are spatial entities with spatial characteristics is to claim a first-order isomorphism between mental and physical structures. Obviously, Kosslyn et al. do not want to make such a claim but it shows how misleading the terminology can be. The cathode ray tube (CRT) analogy in the simulation seems equally misleading in that the CRT contains no information that is not present elsewhere in the system. One can unplug the CRT, and the processing would suffer no major consequences. It is hard to argue that images are functionally distinct representations of knowledge when they are equated with CRTs that can be unplugged without causing any alterations in the processing.

Terminological confusions are in no way limited to definitions of imagery; they also arise in the propositional-analogue debate. The main problem is with the term "propositional" although some confusion surrounds the term "analogue" as well, because there are now three quite different definitions of it in use (cf. Pylyshyn 1979; and "Cognition and Computation" BBS 3(1) 1980).

The proposition is one case out of many in which psychology borrows unexamined constructs from other disciplines (see Keenan 1978). As a construct in logic and the philosophy of language, the proposition was defined as an abstract unit of meaning (i.e., capable of taking a truth-value) which has the form of a relational structure. Psychologists have borrowed this definition *in toto* except that the property of taking a truth-value has little significance because psycho-

logical theories of meaning are rarely based on a concept of truth [see Fodor: "Methodological Solipsism" *BBS* 3(1) 1980]. Now, to say propositions are abstract is not to say anything that would differentiate them from any other type of mental representation; all mental representations are abstract in that they do not share, in a first-order isomorphism, the physical characteristics of the objects and events from which they derive. Also, to say the form is relational is to place virtually no restrictions on the types of information that can be represented. Certainly, spatial information – be it topological or Euclidean – is not excluded by such a representation. Kosslyn et al. argue that propositional representations merely *describe* objects and events and cannot *depict* them as imaginal representations can. But, unless they are arguing for first-order isomorphism between mental and physical structures, the distinction makes no sense; all mental representations are more or less descriptive. What is required, then, to distinguish between images and propositional representations is a more rigorous formulation of propositions such that there is some information that they cannot represent.

Another reason for the continued controversy over imagery is that theorists may take different perspectives on the topic depending on their field of training. Thus, a person who has been attracted to phenomenalistic theories of perception might be more inclined to argue for the unique role of imagery than would a logician or computer scientist. There may also be genetically based differences in cognitive styles among subjects (and theorists) which further contribute to the controversy. In fact, recent research suggests that there are quite significant variations between individuals in the use of imagery which have important consequences for problem-solving behavior (MacLeod, Hunt, and Mathews 1978).

Whatever the reasons for the controversy, the polar positions taken by theorists to date may ultimately be linked by a better understanding of the dimensions of variation in mental processes. Specifically, we need to consider mental processes in terms of the degree to which they invoke operations that relate to the *continuous* perceptual dimensions and reference systems of our environment – space, brightness, time, and the like. For example, the highly studied three-term series problem may be solved with (Huttenlocher 1968) or without (Clark 1969) the use of spatial reference systems (imagery). But this example of dichotomous processing needs to be expanded to capture the view of a continuous dimension between simple logical operations on abstract symbols on one end and imagery processes that invoke operations related to space perception on the other end. This expansion may take the form of including topological reference systems as intermediate between the instantiation of a Euclidean system, such as that involved in most direct perception, and abstract logical operations that involve no perceptual knowledge. Of course, the type of process used may be a joint function of the individual and the demands of the task.

This is not the place to elaborate on a dimensional theory of mental imagery. We only wish to point out that there is already adequate support for such an approach, that it may have considerable practical utility for understanding between- and within-individual differences, and that it may help resolve some of the current theoretical controversy over imagery.

***by* R. Duncan Luce**

Department of Psychology and Social Relations, Harvard University, Cambridge, Mass. 02138

A conceptual, an experimental, and a modeling question about imagery research. Three questions arise as I read this paper:[1] what is the function for us (and presumably other animals) of imaging; are some of the experiments dangerously subjective; and exactly what is involved in the modeling?

Apparently the existence and visual character of imagery are doubted by some scientists. Personally, I have no more trouble with images than I do with dreams, although I agree that the existence of both is difficult to demonstrate convincingly to a skeptic – but then so is the existence of anything except, apparently, thought. For me the issue is neither existence nor visualness, but rather the function and properties of imaging. To be sure, it is unlikely that one can successfully defend any statement of function, but I do find it reassuring to know of at least one useful role that imaging might play in the ongoing life of an organism. I rather doubt that such an elaborate mechanism is either an epiphenomenon or exists so that we are able to answer from memory whether or not a particular dog's ear is pointed, although clearly it can in fact be used for that purpose. It strikes me as far more likely that the whole mechanism is there to aid in dealing with our movements about an environment that we sense in great part visually. My guess is that it is an integral part of our system for visual perception, and its function is to help us decide whether or not parts or all of a visual scene are already known to us. All of the mechanisms of image transformation, so cleverly studied by these authors and others, are there, I suspect, in order to achieve maximally good matches between an image drawn from memory and the current visual display. My view contrasts somewhat with the position of Kosslyn et al. that "there are two sets of processes that we feel . . . are *not* shared with perception. . . . These are the image construction and image transformation processes." Since I find their position surprising and mine plausible, I hope they will point out exactly which data decide between the two views.

Turning to the experiments, a number of them exhibit a difficulty that makes me uneasy, namely, that the subject tells the experimenter when one or another condition is met and the time is measured for that to come about, but the experimenter has no way of verifying that the condition is really met. To cite just one example, in section 1.2 an experiment is described in which the subject presses a button when an image drawn from memory is complete. What assurance is there that the image is in fact complete when the button is pushed rather than, say, 100 msec later, or that it was not complete some time earlier? Or, what is worse, what if it is complete for some subjects and not others? Put another way, such experiments (as with many other nonlinguistic ones in cognitive psychology) cannot be conducted with animal subjects, not because the animals necessarily lack the phenomenon, as with language, but because the design does not afford an objective test that the condition is met. This is not inherent to the area. For example, the rotation experiments end up with a response that is either correct or incorrect, and it should be possible to adapt them to some highly visual animal such as a pigeon.

The modeling has me somewhat confused. At times it appears to be assumed that the information about an image is stored in discrete-valued polar coordinates, which then leads naturally to a CRT (cathode ray tube)-like display which becomes increasingly coarse as one moves from the origin to the periphery. Such a form of storage clearly facilitates radial changes of scale and rotations. However, at other times it sounds as if the information is recovered in the form of a linear matrix, much like a standard CRT, which would mean variations in coarseness by rows and by columns, but not in polar coordinates. Such a display facilitates linear transformations in either the horizontal or vertical directions, but it is not especially congenial to radial expansions or rotations. Were we dealing with continuous representations, then there would be no issue of where the coarseness occurs and the transformation from one form to the other is well known, if slightly complex. But with discrete data, especially if really moderately crude (as the data seem to suggest), no simple conversion is possible. I do not sense exactly how the authors decide which representation to use in accounting for any particular experiment or how much time they allot in going from the polar coordinates of the store to a linear CRT. It is also perhaps worth noting that the polar coordinate representation entails a fixed origin, presumably the point of fixation at the time the image was formed, and that to change from one origin to another involves considerable computation and some change in which regions receive a coarser representation. One wonders if the existence of an origin in the representation could not be demonstrated by rotating a display about several different points, one of which corresponds to the fixation point used in forming the original image. Presumably, all operations on the image are faster when the origin of the display and the image are the same than when they differ.

Finally, it should be noted that the CRT analogy is not the prevailing view about the nature of visual perception. Probably the dominant view, which is somewhat supported by the existence of cells that are

selectively responsive to the orientation of a bar and by various perceptual phenomena and illusions, is that the visual system engages in some sort of Fourier-like analysis. If there is any truth to that idea and if one takes seriously the conjecture that imagery is an integral part of perception, then one cannot but question the adequacy of the CRT analogy. If I were modeling imagery, I would feel some disquiet about that analogy.

NOTE

1. I have also had the opportunity to read the penultimate draft of Kosslyn's *Image and mind* (in press), which treats many of the same issues and experiments in greater detail.

***by* Thomas P. Moran**

Xerox Palo Alto Research Center, Palo Alto, California 94304

The imprecision of mental imagery. The most interesting thing that Kosslyn and his colleagues have done is to expose the "imagist" position – that a visual image really is a picture (to me, at least, an array of points is a picture). I would like to make a few technical remarks about their "analog" imagery model vis-à-vis "propositional" models, since the formulation of a precise model/theory is the central scientific enterprise.

First, some basics. A *representation* consists of a *data structure* plus a *set of operations.* It makes no sense to talk of the data structure alone, since it is only the operations that specify how the knowledge encoded in the data structure is accessed and how that knowledge can be transformed. (I believe that much confusion in debates about representations can be attributed to talking only about data structures, leaving the operations to the diverse imaginations of the parties involved.)

A given range of behavior, such as a body of empirical data, can be accounted for by many different representations – the space of representations is very rich – where the representations differ by trading off effects between the data structures and operations (e.g. encoding the same knowledge in different ways within data structures and thus requiring different operations). I am skeptical (along with Anderson 1978) that cognitive psychology can zero in on *a* representation for mental imagery; I expect that it can only specify a behavioral equivalence class. A clear example of this is Kosslyn et al.'s use of polar coordinates in their "image files"; deciding between this and Cartesian coordinates is well below the threshold of what we can distinguish empirically. At best we can hope to chip away at the behavioral class, and a major part of this will have to be done by considering how imagery fits in with other cognitive processes.

The centerpiece of the Kosslyn et al. model is the two-dimensional image array. According to the arguments above, this data structure by itself should not be taken too seriously, since other data structures, with different operators, are behaviorally equivalent. However, the image array (beginning with the cathode ray tube analogy) is the driving theme of their research. Why do they find it so compelling? I think it boils down to the simple association of this array with the experience of imaging. This seems to me to be a fairly weak argument for the array, since there are a variety of plausible sources for the experience. For example, in my thesis (Moran 1973a) the visual image was modeled as sets of propositions in short-term memory. The simplest identification of the experience of imagery in that model is with the visual propositions in short-term memory, where the vividness of the image has to do with the concreteness of the propositions. In fact, I presented a series of drawings (not a part of the model, of course) showing the states of the image as the subject was scanning over it. Now, it is completely unclear to me why the imagery in this model is any more "epiphenomenal" than in the Kosslyn et al. model.

Kosslyn et al. begin by making a distinction between "depict" and "describe," where a data structure is a depiction if the parts of the depiction represent parts of the depicted object. Technically, this distinction cannot be maintained. There is a whole space of propositional data structures (the artificial intelligence literature is full of them) that *both* depict and describe, from those that mimic a spatial array to those that are very abstract nonspatial descriptions. Of course, any candidate data structure for visual imagery will contain spatial relations of some kind. Depict/describe is not a fruitful distinction.

Kosslyn et al. would have us believe that the choice is between an analog (arraylike) data structure and a totally unconstrained (abstract) propositional data structure. Any other data structure is "unnatural" in that it is contrived to fit the data, the argument being that there is no motivation to constrain the types of propositions in a data structure. The answer to this comes from considering the *operations* on the data structure. Propositions are generated by operations. A generating operation can create only limited types of propositions, since only limited knowledge can be embedded in an operation (e.g. the knowledge about how to recognize certain spatial relations). Another argument by Kosslyn et al. is that attributes such as shape, size, and orientation are "intrinsic" only to a arraylike data structure. They use as an illustration Anderson's (1978) straw-man model of mental rotation, asserting that its "orientation parameter" is totally arbitrary. But this model is a particularly poor candidate for an imagery representation (as Anderson admits). A more reasonable representation would encode the orientation of objects using spatial relations (like "up" or "diagonally up right") and would have operations for transforming these descriptors. Again, it is the generating operations that determine what is "intrinsic" to a propositional representation.

To bring this line of argument to a head, let me assert that, concerning visual imagery, the task before cognitive psychology is in the form of a puzzle: specify the class of representations that explains the available data on visual imagery. This class of representations is to be found by searching in the space of propositional representations. I simply don't see where else one would look. (Array-based representations are in this space.)

An imagery representation would explain what can and can't be imaged, what images can and can't be used for, and how imagery relates to other kinds of human internal representations. The most interesting feature of visual imagery is its imprecision as compared to direct perception. A representation for imagery should explain the nature of this imprecision. There are several possible sources for this in propositional representations. Almost all representations will be constrained by some sort of memory limitations, resulting in partial descriptions at any point in time. The precision of array representations depends on grain size. More abstract descriptions can suffer from incompleteness, abstractness, context sensitivity, and so on.

An array-based representation, such as Kosslyn et al.'s, is surely one possibility to be considered. Taking this candidate seriously will mean facing several questions, which involve the specifications of the design parameters of such a system. What is the grain of the array? A very fine array (say, 500x500 cells) poses the problem of why imagery isn't precise. For example, why can't people *visually* do a scaling task such as the following: imagine a 4-5-6 triangle; is it obtuse or acute? A coarse array lends imprecision, but is it qualitatively the right kind? For example, there are technical difficulties in doing rotations in a coarse square matrix. What is the cell configuration of the array: square, rectangular, hexagonal, circular? What is contained in the cells of the array: bits, intensity values, sets of properties, full-blown descriptions? This last one suggests an interesting candidate: a coarse array is used temporarily to help spatially organize description fragments until operators can come along and build higher-level descriptions. Shouldn't the array be three dimensional? Shepard and Metzler (1971) showed that mental rotation could be done *in depth* as well as in the picture plane. The more detailed investigation by Just and Carpenter (1976) suggests that there is some sort of mental operation that "rotates" 3-d figure fragments in about 45° increments. Perhaps a set of arrays with different configurations is needed. It would seem that developing an array-based representation could lead to a very baroque model indeed!

A final comment: the Kosslyn et al. model seems odd from a systems point of view. The visual system extracts visual features, such as edges, from the retinal array and passes these features to the higher-level cognitive system. What seems strange is to postulate another array in which the same kinds of operations, such as edge detection, must be done all over again.

by **Ulric Neisser**
Department of Psychology, Cornell University, Ithaca, N.Y. 14853

Images, models, and human nature. Why does the theory suggested here strike the reader as clever rather than insightful, as cute model making rather than serious psychology? I think it is because the thinking of Kosslyn and his collaborators is completely detached from everything we know about human nature or about perception, thinking, and the nervous system. Like much contemporary work in "information processing," it attempts to "account for" a sharply restricted body of experimental results (usually reaction latencies) by relating it to an equally restricted class of models (usually computer programs or something similar). The effect is often as if the baby had been discarded and only the bathwater remained.

We know that mental imagery, especially visual imagery, has something in common with perception. (An individual whose mental experience did not resemble perceiving at all would not say "I have an image.") This suggests that a theory of imagery should bear some relation to a theory of perception, at least to the point of indicating how they are similar and how they are different. Unfortunately, Kosslyn et al. are uninterested in perception; they may not even know that it is problematic. Perhaps they believe that we perceive by "inspecting" a "visual buffer" like the one postulated to explain mental imagery. But what could appear on such a buffer during perception? There are only two possibilities, and neither will do. Is it the momentary retinal image? If so, how could the perceiver deal successfully with head and eye motion, changing image size and shape, meaning and recognition, binocularity, coordination of vision with other sensory information? All these would have to be handled *after* the buffer, by operations that Kosslyn et al. do not specify; whatever the operations, they should be available for operating on images as well! Perhaps, on the other hand, a thoroughly processed "percept" appears on the buffer when we see. In that case, the buffer display must have all the properties of phenomenal experience. But this is impossible: our experience is of ourselves moving bodily through an environment of solid objects, and this cannot be duplicated on a visible screen. The notion that some inner eye examines "percepts" when we see simply will not work. (I have argued this point elsewhere; Neisser 1976.) Information-processing theorists are tempted to that notion by analogy with TV screens and cathode ray tubes, but they forget that it is the *whole person* (not some single processing stage) who examines those real displays, using the same perceptual systems that function in more tangible environments.

In arguing that the image buffer model cannot be taken seriously, I do not wish to align myself with all of Kosslyn's critics. Those who insist that mental processes must be discrete and propositional, because that would make them easy to model, deserve even shorter shrift. Processes in the brain are chemical and electrical rather than digital; conscious experience is of a continuing person in a continuous world. To insist that nevertheless there must always be an intermediate level that requires digital or propositional analysis borders on the bizarre. Apparently the advocates of the "propositional" view are simply unable to imagine any structured system other than the digital computer, and any mode of scientific discourse other than the invention of models.

I believe that most of the phenomena reported by Kosslyn's group are "real": the subjects who make odd speeded judgments are genuinely reporting on some aspect of their mental states. But mental states have many aspects, as the early introspectionists discovered; moreover, new ones are easily brought into being by instructions. Making judgments based on remembered information *may* seem like looking at a screen, but it need not. Chronometric results similar to those reported by Kosslyn and his collaborators appear in entirely nonvisual imagining, as when one recalls steps in mathematical proofs that differ in degree of "obviousness" (Hirst 1976) or makes judgments of "goodness" (Friedman 1978). It is probable that similar data could be obtained from congenitally blind subjects (cf. Jonides, Kahn, and Rozin 1975). Such findings offer further arguments against the hypothesis that imagination involves an inner screen where one part of the mind shows pictures to another part. But they do *not* prove that images are epiphenomenal, or that demand characteristics are all powerful, or that the propositionalists are right. The human mind is more continuous, more subtle, and more natural than any of these groups of theorists seem willing to admit, and the problems of cognitive psychology are more difficult than they suppose.

by **Allan Paivio**
Department of Psychology, The University of Western Ontario, London, Ontario, Canada N6A 5C2

Computational versus operational approaches to imagery. Kosslyn's model and theory are explanatory in the same sense that all current abstract cognitive theories based on simulation or more abstract descriptions of a data domain are. The approach is not explanatory or demystifying in certain other senses, however, such as the developmental (learning or other) origins of imagery, the stimulus conditions that control imagery (e.g., the role of verbal cues versus nonverbal ones), and so on. Moreover, there is no theory of the image that explains the elementary units of the model. They are simply taken for granted as descriptive (labeled) primitives. For example, the description of chairs is a factual one, in which "cushion" is one primitive element. As in any componential theory, cushion presumably could be further decomposed into smaller units. In Kosslyn's approach, these units are not based on features but seem to be Gestalt entities of some kind. However, there is no theory of how the entities get there in the first place or what their ultimate nature might be like.

The title struck me as rather ironical. The implication is that imagery has been demystified by a computer simulation model (since that is the focus of the article). In my view, the only thing that demystifies anything is factual information. Kosslyn has certainly contributed more than his share to demystification in such terms. However, demystification also occurred in a big way in the 1960s, with the operational approach to the imagery construct through my own work and that of Bower and others. It remains to be seen whether computer simulation approaches actually contribute to our understanding of the psychological problems associated with imagery, or simply give the illusion of doing so by couching them in terms of the new computer metaphor.

by **Zenon Pylyshyn**
*Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2 **

Imagery theory: not mysterious – just wrong. The Kosslyn et al. paper provides a particularly useful summary and analysis of recent research and theorizing about the mental operations involved in imagery. By placing their theoretical position in a broader perspective the authors seek to clarify some of the theoretical disputes that have divided students of cognition in recent years. Unfortunately, however, the real theoretical issues are not at all as Kosslyn et al. have portrayed them. The current disagreements are, in fact, at a level quite different from that suggested in their paper. Because the point of view presented in the paper is widely shared, it deserves a much more thorough discussion than is possible in a brief commentary such as this. I have endeavored to elucidate the problem as I see it in two forthcoming papers (Pylyshyn 1980a, 1980b), one of which is to appear in the next issue of this journal. What follows should be viewed primarily as notice that a fuller discussion of the issues will be available shortly. Below I will do little more than present a brief summary of the points developed at length in these papers.

1. What is the objection to the Kosslyn et al. position? Unfortunately Kosslyn et al. have completely misunderstood the nature of the opposition that I and others have maintained to models such as theirs. To saddle us with the belief that "imagery is not a well-formed domain in its own right," or with the belief that images are "epiphenomenal," or that the notion of a mental image is "intolerably vague or logically incoherent," or even that it lacks heuristic value or scientific respectability is absurd. In fact, "mental imagery" is not even a topic to which such ascriptions can be applied until someone gives the notion a theoretical explication – as Kosslyn et al. have attempted to do. When

Present address: Center for Advanced Study in the Behavioral Sciences, Stanford, Calif. 94305

such an account is provided, however, the only question that remains is whether it (the theory, not "imagery") is *true* and *coherent*, not whether it is epiphenomenal or anything else. I have repeatedly argued that particular proposals provided by students of the field are simply false as they stand.

While I have pointed out that experimental results are compatible with a wide range of possible forms of representation, I have never endorsed what Kosslyn et al. refer to as "Rube Goldberg" models consisting of such devices as associative networks. Such models do have certain advantages over ones that appeal to imprecisely specified "analogue media." However, as Kosslyn et al. rightly point out, they are ad hoc contrivances. It should be noted, however, that they are ad hoc in precisely the sense in which the Kosslyn et al. model is itself ad hoc – namely, both are a direct response to the experimental results that they are intended to explain. Neither approach separates and validates the fixed principles (the constants of the theory) independently from the particular procedures adopted to mimic the observed behavior in some particular situation (the empirically estimated parameters of the theory), and both hence fail the degrees-of-freedom accountability that every explanatory theory must face.

Because this issue concerns computational models in general, not just imagery models, it is the main topic of my forthcoming BBS paper. There I argue that information-processing models must be severely constrained, since otherwise (and contrary to what Kosslyn et al. explicitly claim) such models provide an "existence proof" of nothing but the theorist's ability to write large programs. I would be willing to bet the remainder of my sabbatical that no class of theory has ever been entertained that cannot in some way be implemented as a computer program. After all, what is to prevent one from labeling some function "MIND'S EYE" and some data structure "IMAGE" (or, for that matter, "SOUL" with property "ORIGINAL SIN"), so long as the behavior of the overall system can be given an interpretation compatible with some particular sample of experimentally observed behavior?

2. The gap between results and rhetoric. One of the most annoying aspects of discussions about mental imagery is that reliable and elegant experimental results are repeatedly paired with far-reaching and highly unwarranted claims. These claims frequently trade upon the connotations of certain terms (such as "holistic" or "experience"), or else surreptitiously appeal to the metaphorical use of physical terms to refer to mental objects (e.g., "larger," "further," "clearer," "brighter"). The latter can easily lead the unwary to feel that something is actually being explained when all that is happening is that the observations are being metaphorically redescribed. For example Kosslyn et al. suggest that their model not only explains certain reaction-time data, but also accounts for the nature of the "experienced image." In contrast, they contend that, "none of the models of imagery based on artificial intelligence research treats the images that people report experiencing as functional representations." Such a contention is simply nonsense. For one thing, people do not report "experiencing" theoretical entities such as *representations* – they report what the *objects* that they image look like (Hebb 1968 has also made this point). It would be absurd to require that some part of one's model *look like* what subjects report. But though this sort of view is nonsense, it is nonetheless surprisingly prevalent. Almost all "image theorists" other than those whom Kosslyn et al. class as being in "artificial intelligence" assume that the computational view treats images as being "nonfunctional concomitants of mental processing." This assumption is patently false. Such theorists simply refrain from assuming that anything that *represents* a scene must actually *look* like it.

Now of course Kossyln et al. do not posit pictures in the head because that would raise embarrassing questions concerning their ontology. A *picture model* would no longer be a functional one, but would constitute a substantive physical or biological hypothesis. What these authors claim instead is that the "CRT proto model" is a useful heuristic. And so it is (though not for the reasons they seem to believe – *cf.* Pylyshyn 1980b). In the heuristic protomodel we have an actual two-dimensional display which has some specific size, shape, and so on. Now the real theoretical explanation, according to Kosslyn et al., comes from the abstract mathematical properties of the protomodel that are preserved in the symbolic simulation. But notice that after this unexceptionable introduction to their position (section 2) we find that it is still only the protomodel that continues to carry the explanatory burden. That is because it is only when you have a *real* distance, that the time taken to traverse it *must* vary with the magnitude of the distance, and it is only when you have a *real* display, that details *must* become blurred when the size of the projected figure becomes smaller keeping resolution fixed. Now of course we can invent data structures and programs to mimic such phenomena. For any result of this kind, inventing mimicking procedures is a simple matter. It becomes more difficult in practice if we attempt to cover a large variety of such results (providing, of course, that these results are not simply variants of the same phenomenon – which, in the cases of "mental scanning" experiments is a very big proviso). But for a model merely to exhibit the same behavior as found in a number of experiments is by itself of marginal theoretical interest. What still remains is to say *why* the model behaves the way it does, and to do this without making substantive reference to the protomodel. In other words we must say *why* it takes longer to "scan" *representations* of greater distances, and we must do so in the same principled way that we do when we say why it takes longer to scan physically greater distances (as in the protomodel). In the latter case the equation, "distance = speed × time" is a universal inviolable law of nature. But there is no law of nature that says it must take longer to go from a *representation A* to a *representation B* when *A* and *B* merely *represent* locations that are further apart in the world. Nor can there be such a universal law, since my pocket calculator would be a counterexample.

The only way out of the dilemma is to sacrifice the intuitiveness of the description that applies only to the protomodel (i.e., the one using terms like "further" or "bigger") and to be more explicit and precise about what is being claimed about the mental mechanism involved in imaginal scanning. What one would have to claim, for example, is that there is a mode of cognitive processing (viz, the imaginal mode) in which it is indeed an inviolable property of the system (for unspecified reasons having to do with the structure of the brain) such that one cannot access information about location *Y* after accessing information about location *X* without processing information about (or somehow obtaining access through) locations X_1, X_2, . . . X_i . . . for all locations X_i that lie between *X* and *Y* in the world being represented. While this story does not give us a principle for *why* this should be so (as the protomodel did), it nonetheless makes an important empirical claim. Unfortunately it is a claim that is very likely to be false, for reasons I discuss in the papers referred to earlier, which I shall outline very briefly below.

3. Fixed constraints versus tacit knowledge. I have argued that the most fundamental distinction that a cognitive model must take concerns whether a particular putative function is a basic biologically determined operation (and hence a fixed capacity of the system), or whether it is determined by symbolically encoded rules and representations – that is, by such things as beliefs and goals. This dichotomy corresponds to the distinction made in computer science between functional architecture (or the underlying "virtual machine") and symbolic computation. If an observed regularity arises from certain beliefs or goals, it tells us nothing about fixed mental capacities. These regularities can often be radically altered by merely varying such things as instructions. I have referred to such functions as being "cognitively penetrable." (Note that, contrary to what Kosslyn et al. allege, the point of this particular distinction has nothing to do with questions of reduction. No precise model can fail to take a stand – albeit implicitly – on this issue. Indeed, this is precisely what Kosslyn et al. do when they refer to the "surface display" as having certain intrinsic properties such as being "spatial.")

To ensure that the results discussed by Kosslyn et al. (e.g., time for mental scanning as a function of distance, time to report details on an image as a function of its subjective size) contribute to our understanding of properties of mind (as Kosslyn et al. claim) – as opposed to, say, telling us what the subject *knows* or what he or she takes to be the task – it must be the case that the reported functions are cognitively impenetrable. In other words, the functions must *always* apply when imagery is being used. But a little introspection should convince one that this is not the case: that one *can* in fact change such functions.

You can make your attention jump from place to place just as easily as you can change its "scanning speed." In fact you can do almost anything you wish with your "mental scanner."

To see that this does not contradict the results that Kosslyn et al. report, we first need to distinguish between two different tasks, both of which are compatible with the instructions given to subjects in all such experiments. Task (1) is to *use your image* in answering certain questions (say, whether a named place is on an imagined map). Task (2) is to *imagine yourself actually seeing* certain physically possible events taking place (e.g., imagine glancing from *A* to *B*, observing a moving spot, etc.). If a subject has tacit knowledge of various properties of such events, or of certain principles by which these events unfold (e.g., that "distance = speed × time"), then to do task (2) properly the subject *must* use this knowledge – otherwise he would be imagining that he was viewing an event he *knew* to be progressing incorrectly, thus violating the task description. But he *need not* use this knowledge to do task (1). The use of such knowledge is, to use Newell and Simon's (1972) phrase, part of the "task demands" of task (2). Notice that this is very different from the kinds of "demand factors" that Kosslyn et al. discuss, and their counterarguments are irrelevant to the present point. For what needs to be established is not whether subjects are second guessing the experimenter or being overly cooperative, but whether, say, the linear relation observed between time and imagined distance is a *necessary* consequence of doing task (1).

In doing task (2) subjects can use *any* facts, memories, or knowledge of general principles that they may think of, and in any combination, to construct a sequence of representations (of an undetermined nature) that correspond to how they believe the event would have proceeded. They need not be confined to one of the few "demand factors" that Kosslyn et al. discuss. Furthermore, even if subjects *are* explicitly asked to do only task (1) they may still prefer to solve the problem by doing task (2) – out of habit or for any number of reasons that have nothing to do with the constraints imposed by fixed properties of the functional architecture (i.e., of the "surface display"). Therefore, the crucial question is whether there are *any* circumstances under which one can be reasonably sure that imagery is being used and yet functions attributable to intrinsic properties of the "surface display" are not observed. Surely the answer must be that such functions can be altered, if not at will, then at least after some determined effort. For instance, it seems intuitively obvious that one can imagine a map and "notice" lights going on simultaneously in different locations, or the offset of a currently attended light being simultaneous with the onset of another to which attention is "switched" with no distance-dependent delay. Several studies in our laboratory (outlined in Pylyshyn 1980b) have in fact confirmed this intuitive expectation. Furthermore, some of Kosslyn's own work (Kosslyn, Reiser, and Farah, in preparation) shows that the dependence of reaction time on image size can also be eliminated under certain conditions conducive to an interpretation of the task as task (1) (e.g., in which the subject is not tempted to imagine that he or she is actually viewing a "zoom" sequence).

Of course if one does not take the model too literally, or if one is not overly concerned with the explanatory power of the underlying theory, facts such as these need be little more than inducements to add additional ad hoc assumptions. For example, one can add a "blink transform" which serves as a free parameter to patch such deviations, or one can argue that such "cognitive penetrations" can be accommodated simply by adding a more flexible overseeing executive process. As Kosslyn et al. propose, this executive could decide such things as whether to scan or to blink and whether to rotate an image, and if so, at what rate. But it must be recognized that the principal attraction of the "surface display" story – namely its greater constraint and hence greater explanatory force – is lost in the process. The "surface display" in this hybrid model now serves little (if any) function since the explanatory burden now falls upon the executive which must, in turn, appeal to the subject's goals and beliefs in deciding what to do. In fact the executive's powers are such that we could, if we were so inclined, dispense with the display entirely and let the executive itself produce the observed results by generating reaction times directly. Thus Occam's Razor puts the burden of proof on those who insist on appealing to some noncomputational "medium" such as Kosslyn et al.'s "surface display."

To conclude, though the Kosslyn et al. paper does provide a useful summary of their position, it is unfortunately marred by a misunderstanding of the alternative point of view. All I have endeavored to do here is to sketch the outlines of this approach. The interested reader is invited to consult the cited forthcoming papers for more detailed arguments and examples.

***by* Alan Richardson**

Department of Psychology, University of Western Australia, Nedlands, Western Australia 6009, Australia

Conscious and nonconscious imagery. Stephen Kosslyn and his colleagues have presented an outstanding example of systematic research and clear thoughtful reporting. Certain basic facts are not in dispute. For example, when subjects are instructed to perform a mental scanning task, response times are found to be analogous to those obtained when performing the same task, physically, in the external world. Objects, whether manually or physically represented, take more time to locate when they are more distant from each other. What is in dispute is the process postulated to account for these results.

The authors consider some alternative explanations, but little is said about the possibility that different people may employ different processes (strategies) and achieve the same results. Because the terms "image" and "imagery" are so easily misunderstood by the subject when receiving instructions and by the experimenter when receiving reports the possibility becomes of enormous importance that strategies of a nonvisual imagery kind may be used. Computer simulation of a visual scanning process is not warranted until it can be shown that this process is the one requiring simulation.

Consider the following example. "People claim that when asked which is higher off the ground, a horse's knee or the top of its tail, the information becomes apparent only when they construct an image of the beast." The implication is that all people necessarily employ a conscious visual imagery strategy. It is not difficult to show both logically and from what some people claim, that constructing "an image of the beast" is not the only way in which it is possible to arrive at an accurate answer. First, the subject may know the answer and simply give it. This is the imageless thought phenomenon and, of course, begs the question as to the kind of nonconscious cognitive process involved. Second, the subject may report that the solution process is accompanied by a vague awareness of a *tactile-kinaesthetic* spatial layout from which appropriate inferences are drawn. Third, a reasoning process may be reported that seems, to the subject, to involve an internal monologue in which the known relations of a horse's knees to its shoulders and hindquarters, and hence to the top of its tail are considered, and an answer is given. Mixtures of these last two strategies and of visual imagery may be reported by some subjects who wish to check their answer more thoroughly before giving it.

The time taken to "scan" from one object to another in the "island map" experiment may be accounted for by at least one of the alternative strategies described above, and consequently response time cannot be used as behavioural evidence in support of a visual scanning strategy alone. Two kinds of study bear empirical witness to this theoretical conclusion.

In mentally counting the number of corners on a Brooks letter E (Brooks 1968) most people have no difficulty in arriving at the correct answer. For some this result may be achieved by the self-talk reasoning procedure – for example, "I'll start at the top left outside corner; that's one; now to the top right outside corner; that's two," and so on. The more corners to be counted the longer it will take. Other people who show the same time-distance effect may arrive at a solution by a less systematic, but equally effective, route. Careful introspection by these subjects does not reveal the presence of visual imagery or deliberate verbal reasoning but a sequence of slight eye and head movements accompanied by kinaesthetic strain sensations in related muscle groups.

Whether these self-observed (or experimenter-observed) move-

ments reflect an underlying cognitive process having a "causal" influence on solution time is unknown, but the necessity of a visual strategy must be denied. Using a Shepard (1978) mental rotation task Marmor and Zaback (1976) demonstrated that blind subjects produced the same time-angle of rotation results as have been obtained by sighted subjects. If sighted subjects have a preference for representing spatial events in the kinaesthetic-tactile mode, there is little doubt that they can convert information presented in another modality (e.g., the visual) into this mode and achieve solutions to mental rotations or other types of spatial problems.

So far the main aim has been to show that the tasks employed by Kosslyn and his colleagues are capable of being solved by nonvisual strategies. Why should it be assumed that nonvisual strategies are employed by some subjects when the instructions and the reports all imply the use of visual imagery? The answer will do no more than raise a reasonable doubt and suggest the kind of remedy required.

It may seem strange to raise doubts about the meaning of an image when everything written in this paper implies, unequivocably, that imagery, of a conscious quasi-perceptual kind, is under discussion. "On hearing the word the subject was to look for the object on his image" is a sentence that refers to some content to which the subject is asked to attend. However, doubts are raised because we know that some subjects use the language of imagery "as if" it had reference to actual sights and sounds. Such subjects do not believe that what they "see" and "hear" in their imagery is very much like actual seeing and hearing. They believe that they and their experimenters are using the languages metaphorically, and not literally. Sarbin (1972) assumes that the use of imagery terms should be treated metaphorically on all occasions. Again, Galton (1883), Roe (1951) and McKellar (1963), to mention only three researchers, have shown that there are intelligent and well-educated people who report either no awareness of visual imagery or a disbelief in its existence. Neomentalists like Paivio (1975a) often write as if consciously experienced images were the referent of their discourse, yet they deny the functional relevance of this form of imagery.

The first essential in remedying this confusion is to recognise a conceptual and functional distinction between conscious and nonconscious imagery. When this distinction has been made it becomes necessary to construct appropriate operational measures for different aspects of each. So far, these remedies have been applied at the individual difference level but not elsewhere. Here is an example with which to conclude this commentary. In a study by Ashton, McFarland, Walsh, and White (1978) no correlation was obtained between the ability to form vivid visual (conscious) imagery as indexed by scores on the revised Betts test (Sheehan 1967) and ability to perform quickly and accurately on a series of mental rotation tasks (nonconscious imagery). Nevertheless, it was found that instructions to employ consciously experienced imagery in the solution of the mental rotation tasks improved performance among those who could voluntarily produce vivid images but had little or no effect among those who could produce weak imagery only.

by **Charles L. Richman, David B. Mitchell, and J. Steven Reznick**

Department of Psychology, Wake Forest University, Winston-Salem, N.C., 27109; Institute of Child Development, University of Minnesota, Minneapolis, Minn. 55455; Department of Psychology, University of Colorado, Boulder, Colo. 80309

The demands of mental travel: demand characteristics of mental imagery experiments. In less than a decade the creative and resourceful work of Anderson, Bower, Cooper, Kosslyn, Paivio, Pylyshyn, and Shepard has broadened our understanding of human cognition. Although differing in their interpretation of the human capacity to represent knowledge (e.g., Kosslyn versus Pylyshyn), their compelling logic and willingness to debate these differences have had a profound and positive impact on cognitive theory. Over the past several years Kosslyn's imaginative research has revitalized psychologists' interests in mental imagery as an explanatory construct. We have no quarrel with the efforts of Kosslyn with his theoretical framework, but are concerned with some of the evidence used to infer mental imagery; see for example Richman, Mitchell, and Reznick (1979), Mitchell and Richman (in press); and Wilton (1978).

In general, our point is that the results obtained in mental travel experiments are probably best interpreted as reflecting several factors. The factors we have mentioned specifically are the subject's history of visually scanning physical distance, the pretest procedure of requiring subjects to scan near and far cities, the implicit scanning and distance cues presented in the mental imagery instructions, and the willingness of subjects to use this information during testing. Richman et al. (1979) demonstrated that demand characteristics are a potential problem in the mental travel paradigm (See Rosenthal & Rubin: "Interpersonal Expectancy Effects" *BBS* 1(3) 1978). Specifically, we found that the description of a mental travel experiment evokes demand characteristics to replicate the results of an actual experiment. As Kosslyn et al. agree, "it is within people's ability to alter their response times if they are so motivated." They suggest that the real issue is whether demand characteristics *are* responsible for distance effects in image scanning in the particular experimental situations in which we obtain them." The real issue for us is to determine *what* factors are responsible for distance effects in image scanning experiments. We do not conclude, as Kosslyn et al. suggest, that our research undermines their explanation of the distance effects. In fact our conclusion (Richman et al. 1978) is as follows: "Finally, we do not argue for or against the notion of quasi-pictorial images with preserved metric spatial relationships. Rather, the present experiments have demonstrated that the experimental demand interpretation may serve as at least a partial explanation of mental travel" (p. 18). Kosslyn et al. argue that just because demand characteristics *can* affect distance estimation, this does not prove that they *do* affect distance estimation. However, if the mere description of a mental travel experiment is adequate to evoke demand effects, then the burden shifts to proving that demand characteristics are not important.

Kosslyn, Ball, and Reiser (1978) report that subjects who follow imagery usage instructions less than 75% of the time are automatically discarded. This procedure may have served to compound the demand effects by biasing their sample, since it can be argued that an eager-to-please subject would be more likely to report a higher imagery usage estimate than a less cooperative one. In fact, a factor analytic study of imagery tests by DiVesta, Ingersoll, and Sunshine (1971) revealed that introspective reports of imagery loaded most heavily on a social-desirability factor.

To ensure that his data reflect image processing, Kosslyn overinstructs his subjects in the use of imagery. However, we would suggest that explicit imagery instruction is by no means a way of guaranteeing that subjects *solely* engage in such processing. Richman et al. (1979) instructed their subjects to use imagery, yet their subjects' responses were significantly influenced by information not relevant to properties of the physical stimulus; subjects consistently produced longer reaction times for map distances labeled 80 miles than 20 miles, even though such distances were physically equal both on the original stimulus and on maps drawn by the subjects. Another aspect of the scanning instructions used by Kosslyn is that they imply movement, which suggests distance-time relations to subjects. Although Kosslyn et al. argue that it is not necessary to mention physical motion to obtain the distance effect in scanning, the alternative instructions they have used such as "shift attention" and "glance up" imply movement by the "mind's eye," which subjects may well interpret as requiring time. Indeed, the strongest evidence for the power of the scanning instructions comes from a comparison of the second and third experiments reported by Kosslyn et al. (1978): when the scanning instructions were omitted (subjects were asked to indicate whether or not a certain object was on the map), the correlation between distance and reaction time disappeared.

Another class of objections that Kosslyn et al. discussed concerns the subject's knowledge of real world relations and of the experiment proper. They cite recent research that has shown that the distance effects in scanning are obtained for three-dimensional distances, even though these do not require greater scanning in a two-dimensional plane. While they argue that the absence of the analog to physical scanning in a two-dimensional physical surface provides support for spatial distances on images, we suggest that if demand effects are

operating, the same results would be predicted with the depth distances. Thus, one cannot satisfactorily discriminate between the two interpretations. Kosslyn et al. also state that most subjects are *not* aware of the purpose of their scanning experiments, as evidenced by the fact that their subjects rarely deduce the correct hypothesis. Recently, Mitchell and Richman (in press) have replicated this finding – that is, very few subjects provide the correct experimental hypothesis when asked explicitly if they thought a relation between distance and time was being investigated. However, 100% of the subjects indicated that such a relation existed, but thought that the experiment involved more than that. Since many subjects who are in fact aware of the hypothesis may not verbalize it, it is not surprising that removal of the subjects who correctly guessed the hypothesis in the experiments conducted by Kosslyn failed to diminish the time-distance correlations.

Kosslyn et al. also suggest that the pseudoexperiment demand effects found by Richman et al. (1979) were so obvious that the subjects had no choice but to respond in the predicted direction. Kosslyn et al. suggest that in a more complicated design (as in experiment 2, Kosslyn et al. 1978), subjects would not attend to the same variables. However, Mitchell and Richman (in press) found that even when 21 distance comparisons are required, subjects in a nonexperiment produce results identical to the time-distance relationships reported by Kosslyn et al. (1978). The principle suggested by Orne (1962) is that if an effect is found in a nonexperiment, then it cannot be solely attributed to the independent variables in the actual experiment, but must be explained by subjects' awareness of what is desired on their part. Although Kosslyn et al. propose that the presence of demand characteristics in our experiments and in theirs are separate questions, we argue that it is an open question: it is not clear whether the variance is primarily accounted for by demand constraints, by characteristics of the image, or by a combination of both.

Kosslyn et al. state that their research has been guided by four issues, which are diagrammed as nodes on a decision tree (see their Figure 1). As each question was answered, they were able to proceed along the tree's branches to answer questions at subsequent nodes. Unfortunately, the tree is asymmetrical; if the answer to the first question had been that imagery was epiphenomenal, the direction of future research would be unclear. There are two alternatives to this dilemma. A positive answer to the epiphenomenal question would not allow any of the subsequent issues to be addressed. However, no research plan would be self-destructive; the plausible alternative is that the tree is symmetrical, such that all of the decision nodes are available under the epiphenomenal as well as the nonepiphenomenal branch. To the extent that imagery research is contingent on the answer to the first question, the two alternatives assume critical importance. If the first question has not been clearly resolved, is the research motivated by the remaining issues invalidated? Possibly, but we would like to argue that if the tree is viewed as symmetrical, then the research has provided useful information about certain components (e.g., retrieval) involved in imagery *processes*, even though the first issue concerning the *form* of imagery is unresolved.

The support for the contention that imagery is *not* epiphenomenal comes from data on image scanning, image overflow, image inspection, and image transformations. The primary arguments for inferring characteristics of the image are based on reaction times as a function of some changes in the stimulus to be imaged. Thus, longer distances require longer decision times under scanning instructions, subjectively smaller objects overflow sooner than larger ones, properties on subjectively smaller images required more time to be "seen," and reaction times increase as a function of both increasing complexity and angular disparity from standard orientations in imaged stimuli. While the inspection and transformation (i.e., mental rotation) data are consistent with a nonepiphenomenal argument, they seem to speak more to the functions of imagery rather than its form: people use imagery to make decisions about stimuli, and the time to make those decisions varies as if something were being physically rotated or inspected. As Anderson (1978) has suggested, such results can be accounted for by nonpictorial explanations, so that these data favor neither form of representation. The image-scanning data, however, seem to imply structural properties in images, and it is such data that form the backbone for Kosslyn's assertion that mental images preserve metric spatial relations from external stimuli. Unfortunately, Kosslyn's methodology and his interpretations of the data are not flawless. Since the heuristic value of the scanning technique is great, we believe it important to identify methodological problems that lead to faulty interpretations. Theoretical issues cannot be resolved by even the most powerful data when the source of the effect is equivocal.

In conclusion we suggest that distance estimation experiments probably tap several factors. One main factor corresponds to the pictorial properties of the image being scanned. Another factor is the subject's ideas about the relation between scanning time and distance and his or her willingness to use this information during testing. Other factors may include interpretation of instructions and pretest procedures. We believe that it is only by explicating the complexities inherent in the distance estimation procedure and attempting to rule out resulting alternative explanations that we can hope to use this methodology as a tool to demystify mental imagery.

ACKNOWLEDGMENTS

This manuscript was partially supported by U. S. Public Health Service MH 21288-06 from the National Institute of Mental Health and the Wake Forest University Research and Publication Fund. Requests for reprints should be sent to Charles L. Richman, Department of Psychology, Wake Forest University, Winston-Salem, N.C. 27109.

by **Edward Sankowski**

Philosophy Department, Northwestern University, Evanston, Ill. 60201

On demystifying the mental for psychology. Kosslyn et al. are engaged in research that it would profit many philosophers to ponder. I shall touch on only a few aspects of their paper, and I shall present criticism, but let me state that I find much of value in their approach.

Kosslyn et al.'s use of the device of a decision tree is perhaps misleading at certain points. I shall illustrate this with their first issue, whether mental images are epiphenomenal (epiphenomenal, one supposes, in that they would play no functional role in supporting memory or perhaps other cognitive achivevements as well). The picture of a decision tree, unfortunately, apparently presents us with the choice between two somewhat implausible views: (a) imagery is always epiphenomenal, or (b) imagery always plays a functional role in supporting (visual) memory and perhaps certain other cognitive achievements. The more likely third possibility that some instances fall into each of the two categories is apparently left out of account. I say "apparently" because it is just conceivable that when Kosslyn et al. opt for the branch of "not epiphenomenal," they may mean to leave open the possibility that sometimes imagery plays a functional role in memory and sometimes it does not. But their view that imagery is not epiphenomenal is ambiguous in this respect.

The formulation of Kosslyn et al.'s, second issue seems dependent on problematic (at least at present) notions of images "simply retrieved *in toto*" or "constructed from parts." Why need it be either always one or always the other? This is another complaint about use of the decision tree. But this criticism will not be elaborated on here. For here the main problem is not so much that there is a third conjecture, determinate in meaning and more likely than either of the two rivals discussed. The main problem is rather that the meaning of such notions as "simply retrieved *in toto*" or "constructed from parts" seems indeterminate (at least at present) in crucial respects. This criticism is supported in part, for example, by the difficulty in seeing what predictions really differentiate between the two rival hypotheses.

The discussion of the third issue (of whether images are retrieved in "coherent units" or "piecemeal") seems to suffer from a questionable inference. Kosslyn et al. point to certain experiments in which specific kinds of images can be intelligibly said to have been "constructed from organized units." It does not necessarily follow from this that images are in general (outside experimental situations altogether, or even simply in different experimental situations) constructed from units and that we need to explain this.

With respect to their fourth issue (whether images are generated only from "depictive" information) Kosslyn et al. suggest that conceptual information can influence image construction. One wonders if they

would grant that this somewhat strengthens the likelihood that at least some instances of cognitive achievements like visual memory, which a subject might sincerely claim are based on information gained from imagery, are actually based on "more abstract conceptual information." This would seem to require adjustment or at least restatement of the authors' views on the first issue.

Kosslyn et al. describe the relation between "the model and the modeled domain" as "one of analogy," a common conception. I wish to emphasize here that philosophers might well contribute to Kosslyn et al.'s inquiry by exploring the conceptual virtues and limits of the analogy between visual images and displays produced on a cathode ray tube by a computer program operating on stored data. There could fruitfully be a coordinated inquiry, psychologists testing out predictions suggested by the model, and philosophers discussing conceptual aspects of the analogy. It will quickly be obvious that glances by each at results of the other may be useful. I shall explore some related issues a bit further.

Kosslyn et al. cite Orne's warning to psychologists "that many of their experimental results can be attributed to the 'demand characteristics' of the experimental setting. That is, S's may deduce the purpose of the experiment in which they are participating and may manipulate their responses so as to give the experimenter . . . the results they think he wants." Kosslyn et al. attempt to meet various criticims which they interpret in light of this reading of Orne's warning. I would like to suggest a related problem. Even if the authors could meet the objections as they interpret them, it still seems that many of the questions they ask their experimental subjects would leave many people uncertain what to say about their own images, or uncertain whether what they do venture to say is true or false, or whether experimenter instructions have been followed. (E.g., it is easy to imagine this justifiable response: "I am simply unsure whether I have scanned my entire mental image from an initial focal point to another particular point on it.") This need not mean that when a subject obligingly answers such experimenter questions, the subject must be consciously or unconsciously trying to give the experimenter what the experimenter wants, in the sense of obligingly trying to support a particular hypothesis favored by the experimenter. But in answering some of these questions *at all*, the subject may be rightly described as "trying to give the experimenter what he wants" in giving any kind of very definite answer whatsoever. This is *not* necessarily to imply that subjects' verbal responses, even when they emerge from such uncertainty, need be either uninteresting or arbitrary. It is of considerable interest what a subject will say under such circumstances, and it is almost always far from arbitrary. But we must remain sensitive to the possibility (not disastrous in itself but generating problems that need to be confronted) that the experimental situation will produce verbal responses that display only a limited degree of matchup with how people would normally talk about (and otherwise relate to) their mental imagery. Also, the information gained from experimental data on such selfconscious image inspection or image modification may well be of only limited application to nonselfconscious imagery (e.g., imagery associated with some memory).

***by* Roger C. Schank**

Department of Computer Science, Yale University, New Haven, Conn. 06520

AI, imagery, and theories. On reading Kosslyn et al. one gets the feeling of the great difficulty that exists in attempting to set forth on uncharted waters. Why have Kosslyn's theories met with so much criticism? I take this article to be representative of the highest quality work in cognitive science, namely, the attempt to build up a consistent model based upon both experimental and computational evidence. Why then is he accused of being ad hoc and not having a theory?

The answer, I believe, lies on the last page of the target article. Kosslyn states that his model will constantly be extended to accommodate the new data. Traditional philosophers of science argue that such extensions refute his theory. Indeed, they are right insofar as the sense of theory that they are applying is the traditional one.

But, in the days of computers that can test complex models, this notion of theory must be revised. From a computational viewpoint, a good theory is one that produces the correct input-output behavior for the "right" reasons. The issue is, what constitutes a "right" set of reasons?

I would like to suggest that any set of computational procedures that handle a limited set of data cannot be judged right or wrong. Ultimately, the right procedures are the ones that handle the most data.

Clearly, if this is the case, then the issues that Kosslyn has been forced to discuss in this article are entirely misguided. The ultimate test of any theory must be its computational generality, and here we have Kosslyn presenting us with the *only* theory that is both computationally and experimentally based. The burden of proof is on his detractors. Let them come up with a better (i.e. more general or computationally efficient) model that works, and then issues of selection criteria will apply. Until that point is reached, all theories can be claimed to be ad hoc; Kosslyn's is merely the most general ad hoc theory. The fact that such a statement seems to be a contradiction in terms reflects the fact that our terminology and our criteria for evaluating models in cognitive science will have to be revised. To Kosslyn, I say: keep up the good work!

***by* Benny Shanon**

Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel

The image-like and the language-like. The comparison between visual and propositional representations is at the heart of an active debate in contemporary cognitive research. The issue – as presented, for example, by Pylyshyn (1973) – is whether mental images can indeed serve as underlying cognitive representations. On this issue, Pylyshyn (1973) takes a negative stand: he argues that underlying representation is propositional, whereas images, albeit subjectively real, are – from a theoretical point of view – epiphenomenal. It seems to me, however, that one can accept much of Pylyshyn's argument and still accept images as viable cognitive entities. For this, one has to distinguish between underlying ("deep") and working ("high") levels of representation. This approach is taken by Kosslyn et al.: they concede that the underlying data base is propositional, but they also show that high level images can be constructed from it. This new characterization calls, however, for a revision in the formulation of the basic issue of the debate. According to the new characterization, images and propositions pertain to two different epistemological orders; hence they are no longer legitimately comparable. The viable comparison should instead juxtapose the two high-level entities: mental images on the one hand and lexical realizations on the other. Both these entities, note, are assumed to be linked to the same underlying structure, which, indeed, is propositional.

The difference between image-like and quasi-verbal representations is characterized by Kosslyn et al. as rooted in their being depictive and descriptive, respectively. Depiction, I take it, is the property that enables a message to be extracted via a direct consideration of the representation at hand, whereas description is the property that requires two entities (i.e. the representation and its referents) to be compared for the message to be extracted. While I agree that the description-depiction distinction is valid, I believe that it is not sufficient to distinguish between the different cognitive representations. For this, another distinction has to be introduced.

The additional distinction I have in mind is perhaps the source of the natural fascination with the current research of mental imagery. This research suggests that there are mental processes rather like operations conducted outside the conceptual domain proper. Unlike lexical or sentential representations, which are abstract, images are, in some sense, concrete. As already noted, the significance of quasi-verbal representations stems from their relation to other entities; in themselves, however, these representations are empty, and the substrate (form, medium) in which they are couched is of no relevance. In this respect, I would say that quasi-verbal representations are *transparent*. In contrast, imagelike representations are *opaque:* the substrate in which they are couched is relevant, and one that may affect their processing as well as their interpretation [see Fodor: "Methodological Solipsism . . ." *BBS* 3(1) 1980] Thus, whereas the descriptive power of words is not affected by their physical size or shape, the depictive power of images – as is demonstrated by Kosslyn et al. – is affected by such factors. This opacity of images is, indeed, the basis of their

likeness to other, more concrete entities, hence the source of their own relative concreteness.

The concrete properties of images may be defined on several dimensions. Studies of mental rotation (e.g., Shepard and Metzler 1971; Cooper and Shepard 1973) suggest that images are similar to physical bodies, which can be handled in a concrete manner. As Pylyshyn (1979) notes, however, nobody had ever tried to attribute to mental images properties such as mass or torque (let alone, I would add, coarseness or temperature). If an image does not possess all of a physical body's properties, however, it becomes necessary to draw the line of demarcation that differentiates between those features that are thus shared and those that are not. Kosslyn et al. suggest another type of target, to which the opaque qualities of images should be likened, namely, objects of perception. But here, again, the problem of demarcation lurks. Are all the properties associated with visual space relevant? Spatial extension is, but what about brightness, perspective, or parallax? The joint consideration of the two lines of demarcation leads to a rather interesting conclusion. Specifically, visual-perceptual properties offer neither an "upper" bound (e.g., rotation) nor a "lower" bound (e.g., brightness) by which the opacity of mental images may be defined. The burden is still on the cognitive scientist to define the constraints characterizing the set of properties that are indeed relevant in this regard.

So far we have assumed that quasi-verbal entities are transparent whereas only images are opaque. Generally speaking lexical and sentential representations are transparent: one zooms through words to their meaning; indeed, one cannot halt and ignore the meaning. (The Stroop effect is the most noted attestation to the reality of this fact.) Yet, in some instances, verbal entities are also opaque. In these cases, the particular phonological, intonational, or even graphemic structures of words are taken as possessing intrinsic significance of their own. Examples to this effect are poetry, mystical systems (in the Kabbalah, for instance, the individual letters are endowed with independent meaning, and even existence), primary processes (Freud 1958; Noy 1973), and pathological thinking (Arieti 1955).

Descriptive representations, then, may be opaque; can depictive ones be transparent? I think yes. The agents introduced in Minsky's (1977) Society of Mind provide an example. These agents are defined solely in terms of their relation to one another; hence they are transparent; on the other hand, to get their message one has to "see," not to "read," hence they are depictive. The arts offer other examples indicative of the independence of our two distinctions – that between the descriptive and the depictive, and that between the opaque and the transparent. Music, at least when not narrative, is opaque but neither descriptive nor depictive. The works of plastic art are all opaque (how else could one speak of style?), but they may be either descriptive (e.g., figurative art) or depictive (e.g., abstract art, see Kandinsky 1947).

Above I have discussed the properties characterizing the opacity of images; let me, then, comment on another class of such properties, which I don't deem relevant. Specifically, I refer to Pylyshyn's (1979) "Intrinsic properties (e.g. physical, biological) of the [operating] system itself." Two domains are specified here, but the joint reference to them, I believe, is both confusing and confused. The likeness between images and the physical domain is one holding between two levels that (in some schematic ontological hierarchy) do not dominate each other. In contrast, the alleged likeness between images and the biological domain is a concomitant of reductionism: it is a likeness holding between two (ontological) levels, one of which dominates the other. Now, it is unlikely that anyone would propose the class of living organisms as constituting the target to which the opaque properties of mental images are to be likened, and further – given our present state of ignorance – nobody would base the intrinsic properties of mental images on the physical (as distinct from biological) structure of the brain. But then Pylyshyn's (1979) philosophical argument is limited to the biological domain, whereas his empirical argument is limited to the physical one. The intrinsic properties associated with these two domains have totally different senses; hence no conclusions can be drawn from one line of argument to the other. Specifically, Pylyshyn's (1979) theoretical arguments cannot be held against empirical studies of mental rotation, and his own experiments cannot bear out the claims he makes as a theoretician.

In closing, let me note that the present comparison of the image-like and the quasi-verbal does not imply that these are the only possible high level representations. Sensory modalities other than vision may generate other cognitive entities, which may in turn exhibit their own types of opacity. The actual existence of representations of this sort may vary with the biological species in question: lower organisms, for instance, may have olfactorylike images.

ACKNOWLEDGMENTS

I thank Kariel Pardo, Amalia Greenfeld and Meir Perlov for helpful discussions.

by **Peter W. Sheehan**

Department of Psychology, St. Lucia, Queensland, Australia 4067

Metaphor versus reality in the understanding of imagery: the path from function to structure. In arguing that images are analogous to displays generated on a cathode ray tube by a computer program, Kosslyn et al. are careful to assert that we should consider such displays as *quasi*-pictorial. Images, not being objects, simply do not have physical dimensions or exist in any entitylike fashion, but the point is made that it can nevertheless be very useful to talk of concepts like distance and size in relation to images. Images are said to represent these features in a fashion similar to the way they are encoded in the representations that underlie our experience of seeing during perception. Images are experienced, for example, as if we were looking at objects, large or small, or near or far away, and these images can be scanned. The metaphor that is implied by discussing things in this way is useful precisely because by allowing us to talk of scanning and focusing in relation to image processing, it helps us to highlight lawful and hitherto unobserved functional consequences of mental processing that are similar to those that exist when we perceive actual objects. Ultimately, however, Kosslyn and his associates aim to work toward asserting the reality of imagery as a structurally distinctive form of internal representation – distinct, for example, from abstract linguistic representrations that may also store distance and spatial relationships.

Support for the reality of imagery and the structural distinctiveness of the image construct derives ultimately from the extent to which the imagery hypothesis can account for data better than other hypotheses. Kosslyn et al. pay particular attention to several rival hypotheses. They consider, for instance, attempts to explain events in terms of information stored in networks of propositions, and they also examine closely the notion that subjects may be responding to cue demands associated with the imagery test situation. The former account argues for an alternative form of internal abstract representation, while the latter view posits an artifactual explanation of events in terms of the presence of social influence factors attached to the test setting itself. Whereas propositional theorizing has been much debated elsewhere (see Anderson 1978; Pylyshyn 1973), the latter account has not, and for that reason is considered here in some detail. Close analysis of Kosslyn et al.'s position indicates that the logic of the artifact viewpoint has been misrepresented.

Kosslyn et al. show that images may have spatial properties; their results, for example, support imagery predictions by demonstrating that subjects' response latencies are positively associated with actual physical distances on a map, the inference being that people deal with mental image scanning in the same way that they deal with visual scanning of pictures. *Richman et al.* (1979), on the other hand, argue that past results may be due to demand characteristics adhering to Kosslyn's procedures, the concern here being that subjects' responses could be a special case of transfer of training in which knowledge of moving objects and their characteristics comes to influence response latencies in the mental travel test situation. It is important to recognize that the argument of *Richman et al.* [q.v.] is not against the notion of quasi-pictorial images preserving metric spatial relationships; rather, their claim is that "the experimental demand interpretation *may* serve as at least a partial explanation of mental travel." (p. 18, italics mine.)

In their paper, Kosslyn and his associates criticize the work of

Richman et al. by arguing that the subjects in their research were exposed to a preinquiry technique in which the procedures could have suggested to them the hypothesis under study. The point is well taken, for the argument that demand variables might not have operated in the same way in Kosslyn's more complicated studies is a perfectly valid one. But what the research does say and what Kosslyn et al. overlook is that demand cues *could* be responsible for the test data, at least in part; it is not at all true, then, that "the Richman, Mitchell, and Reznick experiment says nothing" and that their conclusion is a "nonsequitur." The logic of alternative explanation is what is at issue here. The data demonstrate that the demand characteristic hypothesis poses a possible alternative explanation of results, thus placing the onus on researchers who adopt Kosslyn's procedures to show that the artifact hypothesis is not, in face, viable, or that it is irrelevant to the test data. Kosslyn et al.'s appeal to evidence gathered in a postexperimental inquiry, though relevant, is not adequate by itself. Data from the postexperimental inquiry procedure, just as with the preinquiry strategy, may reflect the suggestion to subjects of cues that were not operating in the actual study, or they may equally reflect the fact that subjects are engaging in a pact of ignorance in which the best interests of both experimenter and subjects are served if subjects do not admit to the actual hypothesis under test [see also Rosenthal & Rubin: "Interpersonal Expectancy Effects" *BBS* 1(3)1978].

Kosslyn et al. review some alternative theories that might account for the possible artifact and attempt to discount them. The study of artifact, however, is best viewed as a two-stage process in which attention should first be given to locating those cues that are actually implied by the test procedures used and then, once these cues have been isolated, to theorizing about the ways in which demand effects might be mediated, with a view to developing procedures that will modify their potential influence. Like other plausible (and parsimonious) scientific explanations, demand characteristic theorizing requires objective analysis, which may be provided by the study of artifact through the manipulation of cues by instruction (see Singer and Sheehan 1965), or by the application of quasi-control conditions such as role playing. Strategies of these kinds, which basically aim to overcome the essential limitations of verbal report, have yet to be employed with respect to Kosslyn's model. The point is a particularly important one given that Kosslyn's procedures follow the pictorial metaphor so intuitively.

The analogy adopted by Kosslyn is a strongly compelling one, particularly to those who can experience the phenomenon of imagery. It seems especially critical, then, that Kosslyn et al. distinguish, or at least point to the means by which we can distinguish, aspects of their model that reflect mere metaphor from aspects that reflect "substantive theoretical claims" about the reality of imagery. Despite the ingenuity of Kosslyn's empirical work indexing the functional specificity of imagery, the path from function to structure traversed by the model is not very clear. Part of the problem is that the proposed model does not readily allow one to approach questions relating to structure in the same sense that is conveyed by other approaches to the study of imagery. The selective interference paradigm (see Brooks 1967), for instance, draws its predictions directly from the assumption that perception and imagery share the same processing features. Although the critical factor may not be the particular sensory modality of the internal representations that are being manipulated, the interference model is more immediately concerned with the structural similarity of imaging to perceiving.

Partly also with a view to approaching the question of structure more explicitly, Shepard and his associates argue now for a paradigm that more directly contrasts imagery and perception by including a perceptual control condition (see Podgorny and Shepard, 1978). It is true, as Kosslyn et al. say, that a psychological theory ought to specify what the brain can do during the course of cognition without requiring any direct reference to the brain itself, but when the theory in question relies so closely on the argument that images are a structurally distinct form of internal representation, it is disappointing that Kosslyn's computer simulation model, depending as much as it does on the pictorial metaphor, is ill equipped to resolve the issue. The problem, though, is shared to some extent by other accounts of imagery, and the interference model as well as other existing paradigms have their own particular limitations (see Sheehan 1978).

At present, Kosslyn's simulation model presents us with a tool for inquiry into the processes and function of imagery, and that tool works well within the limitations of the model's conceptual overlay. The data that have emerged from its application have added significantly to methodological sophistication in the field and have sharpened debate on issues such as the functional similarity of imagery to perception. It is still not clear, however, whether all the objectives that our tools of inquiry into imagery should serve have been canvassed or acknowledged by the model, given that we are willing to acknowledge the genuineness of imagery experience. What are the limits, for instance, to the phenomena of imagery (or its various functions) that the model is equipped to explore? From our introspections many qualities other than those illustrated by features such as focusing, scanning, and size judgment are evident, and these too require exploration. Consider, for example, the extent to which the structure of imagery (not just its content) might be affected by needs, motivations, and expectations. Kosslyn et al. recognize the relevance of these and other pursuits in their concern about the extent to which the imagery process is penetrable by cognitive factors. In its current state of formulation, however, the model does not adequately delineate structure from function. Locked into its metaphor, the complex realities of mental imagery can all too easily pass us by.

***by* William E. Smythe and Paul A. Kolers**

Department of Psychology, University of Toronto, Toronto, Ont., Canada M5S 1A1

On spatial symbols. By its very nature, imaging depends upon introspection for its description – upon a phenomenology of "the internal eye." But descriptors of imaging do not. The descriptors, the words of language, are derived from interaction with the events acquired through the external eye. Serious problems in the description of imaging occur when the words appropriate to consensually verifiable events are applied uncritically to the products of private experience. In particular, the emphasis upon spatiality in imaging, and especially the proposal that its spatiality implicates a mechanism of imaging, would seem to suffer from this confusion.

The target article of Kosslyn, et al. provides a detailed instance of this flaw. On their view, spatial and pictorial predicates can be applied to imagery not only in the analogical or metaphorical sense derived from experience with physical objects, but because such words are thought to capture important features of the psychological mechanisms underlying imaging. Moreover, Kosslyn et al. consider that a defense of this view is provided by the substantial research program and computer simulation model that it has stimulated them to undertake. But of course no amount of work can by itself prove the correctness of the ideas that stimulated it. Hence what we are left with is the assertion that spatial and pictorial predicates that can be applied to private experiences somehow explain those experiences. There are two serious objections to this line of approach, one formal, a second methodological.

In defense of their arguments, Kosslyn et al. refer to Putnam's (1973) rejection of "reductionism" as a style of psychological explanation. However, any explanation requires a "reduction" of some sort in which aspects of the multiplicity of events are captured by a smaller or more tractable number of variables. If the terms used in an explanation are just as opaque, unique, or numerous as the events to be explained, the account formulated in those terms clearly fails in this important formal requirement of a model or theory. All that is required of an explanation then is that its terms be part of a system of description that applies to a number of phenomena, and not just to the one being explained. Merely applying the language of physical experience to the phenomenon of imagery does not explain it.

What has to be explained is how the imager can represent to himself, by means of his own internal activity and without making use of any external markings or inscriptions, objects or events that can be described pictorially or spatially. To say that a person does this by constructing a kind of mental object that has the spatial and pictorial properties that the words describe is to beg the question rather than to explain the event. How the private event is generated remains a

mystery, on that account.

On our view (Kolers and Smythe, in press), images are better taken as symbols in a personal symbol system than as "analogues" or "propositions," and what is required is an account of how the symbols are generated and manipulated. The simulation model of Kosslyn and his colleagues gives the appearance of being an account because it proposes a systematic description of the events of interest. The authors acknowledge, however, that they are uncertain about which features of their model are part of their theory; if this is the case, the model can hardly be said to exemplify the claim for spatiality of imagery, for perhaps some of the features needed now in the model will ultimately be unnecessary. In an adequate theory, "imagery" will be reduced to some set of primitives that are behaviorally appropriate and whose manipulation will predict other related aspects of behavior. That sort of account does not seem to be present in the proposal of Kosslyn and his colleagues.

Our second objection to the views Kosslyn et al. have put forward has a methodological base; we believe that they have confused representations with their procedures and have taken the former as symptomatic of the latter. As we remarked earlier, the only evidence of imagery is the internal observation. The question of what introspection is and what it does has not been resolved yet, despite long dispute. We do not resolve it here, but point out that the internal eye does not see the procedures or operations of mind, ever, but only some outcome or product of mind's operations. We call "naive phenomenology" the view that introspection tells one much about how one achieves some end, whether it be imaging, perceiving, thinking, remembering, using language, or any other cognitive process. The role of consciousness in naive phenomenology is at least equally muddled. Perhaps the principal notion is that in some way consciousness allows one to come into contact with the physical world; and it is introspection that allows one to assume something about consciousness and about the world. (We may make inferences about the world on the basis of our reaction to it, but we are not cognizing the world, then; only ourselves [see Fodor: *BBS* 3(1) 1980]. We would suggest as an alternative view that consciousness be regarded as a sensing organ for orientation in the world, an organ that monitors the state of the organism, not the state of the external world.

What is required for a fruitful account of these procedures of interaction is a system of symbols whose acquisition and manipulation form the personal symbol system that all people use for assessing and relating to the world. The meaning of our terms of reference, the way in which we come to agree on assignments of meaning, the residues of private meaning that our symbols retain, the skills we acquire for manipulating symbols, and even the relative contribution of pictorial and linguistic descriptors in these symbols constitute a firmer basis for discussing cognitive processes than what is currently available. Taking the personal symbols of imaging as symptoms of the process of image formation would be to take the inscriptions of any symbol as indicative of its construction or its reference – a wholly inappropriate procedure.

by **David L. Waltz**

Coordinated Science Laboratory and Electrical Engineering Department, University of Illinois, Urbana, Ill. 61801

On the function of mental imagery. An adequate psychological theory of mental imagery should not only specify *what* we can do, but also give satisfying answers about *why* we do it. What is the function of mental imagery? Kosslyn et al. do not consider this question, and this lack may help account for the resistance to their assumption of the reality of mental imagery. They *do* attempt to establish that mental imagery is not an epiphenomenon by showing that subjects perform differently when they are instructed to image and when they are not; however, they do not attempt to demonstrate that mental imagery is necessary for cognitive tasks, nor do they give a rationale for its origin. Under these circumstances, it is difficult to see what difference the presence or absence of a theory of mental imagery would make to an overall cognitive theory. I believe that mental imagery *is* an important area of study, and in particular, that an understanding of the phenomena of mental imagery can provide important contributions to a theory of perception. I offer here some arguments in support of the reality of mental imagery and nonpropositional representations. I also discuss some shortcomings I see in the theory and model presented in the article.

Let us start with the assumption (which I hope is noncontroversial) that perception involves both bottom-up (data-driven) and top-down (conceptually driven or hypothesis-driven) processes. Kosslyn et al. feel that the processes of mental imagery are top-down and *not shared with perception.* I suggest instead that processes responsible for mental imagery also take part in the performance of virtually all perceptual tasks, and that "pure" mental imagery results when the top-down processes of perception are exercised in the absence of any data to drive the bottom-up processes (as when one's eyes are closed). To support this point, I will first discuss four centrally important examples of top-down processes in perception which have been suggested by researchers in artificial intelligence (AI). I will refer to them here as "cue interpretation," "hypothetical world," "world maintenance," and "scene description encoding" processes. Like the mental imagery operations that Kosslyn et al. have studied, these four processes require the ability to rotate, translate, and perform perspective transformations on three-dimensional mental items.

Cue interpretation. Kosslyn et al. state that "people do not have to use special processes to construct or transform perceptual 'images': the physical environment and their peripheral visual systems do that for them." This assertion contrasts sharply with the current AI view and experience. Introspectively the perception of objects *seems* to be primarily a bottom-up process. Often, however, both in computer and human vision, only incomplete data are available for items in a scene, as when an object is partially obscured, poorly lit, or seen for only a very short time (Waltz 1979). The major current view in AI (based on difficulties with bottom-up systems) is that *all* perception involves varying proportions of interacting bottom-up and top-down processing, depending upon the particular scene and the current goal of the perceiver. A few perceptual cues may suffice for me to "see" that my wife is in the room of our house where I expect to find her. "Seeing" in this case involves a relatively large amount of top-down image construction with relatively little bottom-up processing. On the other hand, a task such as deciding whether I have cleaned all the food off a pot I am washing involves a much larger portion of bottom-up processing.

Hypothetical worlds. Consider perceptual tasks such as deciding, without moving it, whether an object could fit in a given space (related to the "findspace problem" Sussman 1973 in AI), or imagining how a room would appear with the furniture rearranged or a plant added, or finding the piece that fits in a space in a jigsaw puzzle. Unlike "pure" mental imagery where objects to be operated on come from memory, the objects here may be present in the scene. It could be argued that we do not move an *image* but rather a few properties of each object (e.g. the colors and the arrangement of protrusions and indentations of a picture puzzle piece, not the precise shape and picture pattern), but this may only be true for cases in which the object to be moved is unfamiliar or complex. If the space in the puzzle resembles a profile of a face or the outline of the United States, we have no trouble moving it in toto.

World maintenance. Here I refer to the problem of keeping one's orientation in an environment while one is moving through it and when it cannot be seen directly. Special world maintenance problems arise when one is moving backward while looking forward, when one looks down to tune a car radio while driving, or when one enters a totally dark but familiar room. On a more subtle and pervasive level, there is a need to explain how we maintain the conviction that we are perceiving a stable, detailed, 3-D world, even though our view is constantly changing (via saccades, scans, and body movements) and at any given time we can see in detail only the small portion of our environment toward which our foveae are directed. Minsky (1975) discusses these issues, and suggests that we must appeal to *predictive* top-down processes to explain how it is that our models of our surroundings can be updated as rapidly as they are.

Scene description encoding. Because of the relatively greater encoding efficiency of the brain for *pictorial* as opposed to *linguistic* material (Haber 1970), it may be more economical to convert descrip-

tive text into mental images for storage than to store the text more directly. A similar argument comes out of our experience in constructing a computer model for understanding spatial locative prepositions (Boggess 1978, Waltz and Boggess 1979). Suppose that we wish to encode in a computer program a text passage describing the relations among objects in a scene, for example, "A pencil is on a book, the book is on a desk, and the desk is in a room." Note first that the prepositions (in, on, etc.) must be given more precise internal representations; *on* has very different meanings " a shadow *on* the wall," "the nose *on* your face," and "the book on the desk." To be able to form a propositional representation we must (a) define an exhaustive, mutually exclusive set of preposition definitions (e.g. on1, on2, . . .), and (b) construct procedures for picking the appropriate definitions in each case. Second, to answer questions about the relations that hold between various scene items (e.g. "Is the pencil on the desk?") we have to construct (potentially long) deductive chains of rules such as "If A is ON1 B, and B is ON2 C, then A is ON1 C" to see whether any possible meaning of *on* applies in the scene described. Reasoning by chaining such rules can inherently lead to combinatorial explosion, and a complete set of such rules is at least very difficult and perhaps impossible to construct. In contrast, we have shown that it is rather easy to program a system to construct a unified model of a scene consisting of the approximate location, size, and orientation of each object in a global coordinate system, and then to query the model directly. The unified model thus resembles a mental image and allows one to answer questions such as, "Is the pencil in the room?" by making a single comparison of the pencil's coordinates with the room's coordinates in the unified model.

Texts describing events seem even more difficult to handle via propositional representations (such texts are not currently handled by our program). For example, if I were to tell you, "A dog bit a mailman," and then to ask where (on his body) the mailman was probably bitten, you would probably answer, "On the leg;" such an answer may plausibly be a default value in a BITING script (Schank and Abelson 1977) or part of general knowledge. However, if I also tell you that the dog was a doberman or that the mailman was lying on the ground at the time, then different parts of the mailman's body become likely biting sites. The processes we use to answer this type of question seem to me to be unrelated to scripts and intimately related to the processes of mental imagery; I cannot answer such questions myself without imaging. I conjecture that imagelike storage is extremely compact and allows one to encode in one "chunk" not only all the information that would require several chunks if the text were stored directly, but a great many other inferred relations among the text objects as well. Such imagery encoding may also allow one to check for consistency and completeness of the text description. Along these lines, experts in memorization have long used visual imagery to aid retrieval of items (Luria 1968, Bower 1970). It would be interesting to see Kosslyn et al. explore further the relation between mental images and memory.

At this point let me draw some conclusions from all this.

1. I have argued above for the role of top-down constructive processes in ordinary perception. Introspectively, however, it is rarely evident that some portions of a scene are constructed by external data-driven processes, while others are constructed by internal hypothesis-driven processes. It thus becomes reasonable that mental images could be (or at least could *seem*) very real indeed.

There seems to be no difficulty in supposing that higher levels of the visual cortex could be directly driven by memory; neurophysiology seems to provide some evidence for the presence of appropriate connections (Pribram 1969).

2. Mental rotation, translation, and perspective transformations of images from memory must be performed for all the top-down processes mentioned above in order to fit the items in appropriately with data-driven image portions and to verify hypotheses. This strongly suggests that analoglike rather than propositionlike data structures are more natural for representing the scene portions generated by top-down processes. Related arguments are made by Kosslyn et al.

3. If we take seriously the idea that mental imagery is the top-down component of ordinary perception, there are interesting implications for a theory of perception. In particular, some kinds of items will be easy to image or will be expected in context and some will not; one could predict that easy-to-image and contextually appropriate "chunks" are likelier to be filled in by top-down processes, whereas harder-to-image, unexpected objects, or "nonchunked" items would generally be inserted by data-driven processes. We would probably be more prone to err about the presence, absence, or nature of the readily inserted. Treyens and Brewer (1978) have presented some evidence for these kinds of expectation effects.

While most of this commentary is meant to supplement and strengthen the authors' position and research, I also feel there are some fundamental weaknesses in their theory and model. First, the CRT analogy is inherently two dimensional, whereas imagery and perception are inherently three dimensional (or possibly "2½ dimensional": that is, 2-D images with depth and surface orientation information appended; Marr 1978). The experimental results cited in Section 3.1 contrasting perceptual and image "scanning" times support the greater-than-2-D nature of imagery. Surely the difference between the theory and model in this case is important and will probably lead to future difficulties if the model is not modified. In particular, I believe that the Shepard and Metzler (1971) 3-D rotation experiments could not easily be simulated in the model.

Second, the notion of the "mind's eye" that interprets the display is given very little attention. Kosslyn et al. state that they expect the "mind's eye" functions to be shared with those of perception (no argument) and that its function is to "detect patterns in a spatial representation at some level of the visual system." It seems to me that at the level of the CRT-display/mental-image, the "mind's eye" deals with images that have already been organized into nameable units and objects, and that no pattern detection will be going on. I suspect pattern detection is a much lower-level process, important in integrating region fragments into coherent objects. In my view, the "mind's eye" has access to perceptual information at the point where data-driven processes have already carried out the organization of raw data into objects, and where only rather large units can be inserted by top-down processes. The function of the "mind's eye" seems likely to be involved with selective attention to and interpretation of particular image items, item properties, and relations between items.

A third problem involves a confusion about what the "mind's eye" can do; for example, consider the use of the term "scanning" in this article. Does the "mind's eye" scan or does the display move or both? How do these movements relate to scanning by the eye? Surely we do not want to postulate a "mind's eye" with all the properties of a complete perceptual system – this would violate the reductionist assumption that each component of the imagery system is simpler than the entire system. In Dennett's (1978) terminology, we do not want to postulate a homunculus within our mental imagery system. While I cannot offer any neat solutions, I feel that it is imperative that the authors characterize in much greater detail the nature and function of the "mind's eye" portion of their theory and model.

Overall, I applaud this work. Rather than being an embarrassing appendage to a theory of perception, mental imagery may well be central to perception, and may offer an invaluable key to the construction of a comprehensive cognitive theory.

Authors' Response

by **Stephen M. Kossyln, Steven Pinker, George E. Smith and Steven P. Shwartz**

Department of Psychology and Social Relations, Harvard University, Cambridge MA 02138; Center for Cognitive Science, MIT, Cambridge MA 02139; Department of Philosophy, Tufts University, Medford Ma 02155; and Department of Computer Science, Yale University, New Haven CT 06520

The how, what, and why of mental imagery

The commentators have raised many insightful and interesting

points, far too many for us to respond at the level of detail prevailing in the commentaries themselves. Thus, we have tried to respond to those that we took to cut most deeply, and we have tried to cover as broad a range of topics as possible, and at the greatest possible depth. Ideally, we would also have liked to explore some of the many interesting ideas that we found to be congenial and supportive of our enterprise, but we have not been able to do so here; this should not be taken as a lack of interest or sympathy on our part, only a lack of space and time.

I. Methodology

Some of the most important criticisms of our project are directed at the validity of our methodology. These criticisms challenge the empirical foundations of our theory, and hence they must be addressed at the outset.

Demand characteristics. HANNAY, PYLYSHYN, RICHARDSON, RICHMAN ET AL.and SHEEHAN have all raised the specter of demand characteristics in one guise or another. These authors worry that our results reflect, at least in part, the subjects' responding to implicit demands in the instructions given to them [See Rosenthal & Rubin: *BBS* 1(3) 1978]. We have four responses to this criticism. First, consider an experiment just completed by Kosslyn, Jolicoeur, and Fliegel: An initial group of subjects was given statements like "A bee has a dark head," and asked to decide whether these statements were true or false. After deciding, they were asked to rate how much they thought they had used imagery in evaluating the statement. Half of the statements were true, and half were false. The data from the true statements allowed us to assemble a set of object-property pairs that reportedly required imagery to evaluate and a set that reportedly did not. Further, the items were selected such that the properties were all localized on an end of the object. These items were then used, with a new group of subjects, in the following experiment: Subjects were asked to image an object and to focus mentally on a particular end of it. Following this, a property was named and the subject was to decide as quickly as possible whether or not the object had that property – without necessarily using imagery. It was stressed that although the subjects should begin with an image and focus on the specified location, they should decide whether or not the property was appropriate as quickly as possible without necessarily using the image. The point of fixation was varied such that on half the trials the valid properties were on the end at which the subject was focusing and on half the trials they were on the opposite end. The items were presented in random order, and in fact the subjects were never told anything about the fact that two kinds of items were mixed, those previously rated to require imagery and those previously rated not to require imagery. The results were straightforward: For the items that had been rated as requiring imagery, less time was needed if the subject was focused initially on the end of the imaged object at which the property was located. For the items rated as not requiring imagery to evaluate, the distance from the point of initial fixation to the property made no difference whatsoever. And in fact subjects reported that they sometimes had to use the image and had to scan to the location of the property. These results cannot be explained by any reasonable demand characteristic account, given that the high and low imagery items were mixed and the subjects were never told to scan. If demand characteristics were operating, we would never have expected the selective effects of distance on only the high-imagery items.

A second point about demand characteristics is in the same vein. A number of our results would be very difficult for subjects to produce intentionally. Consider four examples in the literature: (1) The visual angle of the mind's eye calculated from one group of subjects allowed us to predict the amount of time other subjects would require to scan the longest possible nonoverflowing image of a line. The image-scanning group would require a host of paranormal abilities to deduce the expected results (and in fact the experimenter did not know what results should be obtained from any given subject at the time the experiment was conducted). The original results are reported in Kosslyn (1978a), and a replication is reported in Kosslyn (in press). (2) Similar effects of the subjective size of parts of an image were obtained with first graders, fourth graders, and adults (see Kosslyn 1976b). Given children's notorious lack of awareness of their own cognitive processes, it seems unlikely that the children would infer the hypothesis and regulate their response times accordingly. Further, in this experiment subjects began by evaluating a set of items without being asked to use imagery. After the task, they were asked whether they had spontaneously tended to "look for" the named properties on images. When data from the first graders were examined in terms of reported strategy, effects of size of properties were found only for those subjects who claimed to use imagery spontaneously. This result cannot be interpreted in terms of implicit demands in the instructions, since imagery was never mentioned at all. (3) Kosslyn, et al. (1977) report experiments in which subjects were asked to begin by forming a normal or small image of an object and then were asked whether that object was in fact larger than a second named object. Even though subjects were never told to use their images, decision times were slowed down when a small image was formed initially for items that the theory predicted should require imagery. (4) Pinker and Finke (in press) had subjects scan an image of a 3D display that they had previously rotated mentally to a new orientation. The time they took to scan between pairs of objects suggested that they had mentally rotated the cylinder a constant fraction farther than the task had called for. Interestingly, the same unanticipated distortion was found in other experiments using two other psychophysical measures of imagery fidelity not related to scanning. It is hard to think of a demand account of image scanning that could explain why the same types of distortions would arise when *other* measures are used. Rather, the most parsimonious explanation is that the images themselves were distorted, and that the three dependent measures recorded this distortion. (See also Finke [in press] for further examples of imagery phenomena unlikely to have been deduced or deliberately generated by subjects.)

Third, at least some of our critics, (RICHMAN ET AL. and RICHARDSON) concede that demand characteristics may not be all that is going on in the scanning experiments. These authors focus on scanning in its own right, and suggest that to understand scanning we must consider the role of demand characteristics (see PYLYSHYN as well). We have not studied scanning as something of interest in its own right, but as a means of answering higher-order questions. In particular, we have used scanning as a kind of "tape measure" to discover whether images depict inverval information about spatial extent. As long as demand characteristics do not interfere with this use of scanning, we are not disturbed by the possibility that they are in principle capable of affecting scanning per se.

Fourth, in the present theory we have claimed that some of the properties of images, such as transformation rate, are under strategic control (see Kosslyn, in press). Thus, it should not be surprising that subjects can mimic certain physical properties – say, the effects of angular inertia on rotation rates – should they so desire. After all, one of the purported purposes of imagery is to simulate possible operations on objects in the world (see LUCE, HUNTER, and WALTZ). This sort of response to demand characteristics is not only of possible interest, but is in no way inconsistent with our theory.

The tree. The present research program relied on a "decision tree" to help clarify issues before the model was constructed. Several commentators had difficulty with this approach. RICHMAN ET AL. argued that the tree should be regarded as symmetrical. This misunderstands the nature of the tree. The tree schematizes issues that are raised when a previous issue has been resolved in a particular way. So, for example, it is of little scientific interest how an image is generated if in fact the surface image is epiphenomenal. SANKOWSKI asks about the particular issues themselves. To clarify this, it is important to realize that the tree was constructed within the framework of a "capacities" (or "competence") orientation; we were wondering what sorts of capacities people have, what sorts of things they *can* do – not what they will always do in some situation. If we

demonstrate that surface images do in fact depict information, this must be accounted for by a theory; if images can be generated from units, a theory must have provisions for explaining this. This is not to say that surface images are *always* functional in every task or that images are *always* generated from parts. But the fact that the system is constructed so it can operate in given ways places constraints on a theory.

Introspective data. LUCE, RICHARDSON, SANKOWSKI, and SMITH & KOLERS worry that our data are too subjective, that we take subjects' introspections too seriously. It should be noted that our data are not in the same class as detailed introspective descriptions that require a good deal of interpretation on the part of the introspector. Rather, we have pared away our tasks until the introspections required are rather simple (in most cases only calling for a yes/no decision). Further, these types of data allow us to test conclusions suggested by our introspections. If it seems as if a smaller image is less resolved, for example, more time should be required to inspect it. This technique is far superior to that apparently used by PYLYSHYN and HEIL, who rely on introspection and intuition in making arguments. It is one thing to say that images evince no properties of pictures, but quite another to account for data from more disinterested introspectors (i.e., naive subjects), which suggest the contrary. In any case, it is often possible to get more direct confirmation of the state of a subject's image, and, needless to say, we exploit these opportunities when they arise. For example, Shwartz (1979) confirmed that subjects had rotated their images by a given amount by requiring subjects to template-match their images against an actual "probe" pattern flashed in front of them. Subjects responded more quickly when the probe had the same orientation as their images were supposed to have had strengthening the notion (suggested by the time taken to rotate the image in the first place) that the rotation was performed as instructed. Finally, in some cases it is of course possible that our data are not reliable or valid. But if so, it is the critic's burden to explain why they are so coherent, why so much convergent evidence from different tasks is obtained (see Kosslyn, in press, for a far more detailed review). Essentially, then, our defense is the same as that offered by workers in the field of psychophysical scaling, where the data today often consist of numbers that subjects assign to "magnitudes of sensations": Such numbers *could* be random fabrications, but the fact that they are systematically related to a theory and are so internally coherent belies this possibility.

II. Particulars

A number of commentators questioned particular details of the model and theory, which are discussed further in this section.

The model. Some commentators have erred in their portrayal of details of the model and how they relate to the theory, and some have not understood why the model gives rise to particular predictions. First, HINTON has made three unfounded objections against our model:

1. It is simply not true that we confine the transformation processes to dilation and translation – see Kosslyn (in press) and Shwartz (1979) for detailed discussions of how rotation can be and has been implemented.

2. The use of two separate arrays to account for the effects of image resolution is an option that we discarded immediately after our first implementation (see Kosslyn and Shwartz 1978); at present, an "inverse mapping function" keeps track of the relation between depicted dots and underlying coordinates (the current implementation is an actual function, not a data-structure simulating one, as was described in Kosslyn and Shwartz 1978).

3. The difficulty in reparsing complex images can be explained in our model without introducing any ad hoc mechanisms. As we mentioned in the target article, there is evidence that images can be constructed a portion at a time, with the portions corresponding to the elements of a description if the image is generated from a description. Another body of evidence (see Kosslyn, in press) suggests that imaged material begins to fade as soon as it is generated. Thus the different parts of the image will be at different "fade phases," with the more recently constructed parts displaying the greatest contrast or definition. Since pattern recognition processes are notoriously sensitive to Gestalt laws of grouping, such as similarity and good continuation, any pattern that cuts across portions of the image showing different degrees of activation will be difficult to detect. Therefore, patterns that cut across portions of images constructed at different instants (i.e., the different portions mentioned in the initial description) will be difficult to recognize.

KEENAN & OLSON point out that there is evidence that parts of an image often do not exist until focused on. Our model makes exactly this prediction, given that only a limited number of parts can be maintained in an image at once. In the simulation, if a part is not found at a location, that region is further elaborated and then inspected again. Kosslyn (in press) described several experiments that provide support for this claim. DE VEGA argues that the interpretive procedures cannot be independent of inspection procedures because parts can often be generated when they are needed. In the model, distinct processes are used in interpreting spatial patterns and in converting long-term memory encodings into spatial displays. The mere fact that two processes invoke each other or work together does not mean that they are not conceptually distinct.

We have claimed that the model allows us to make predictions and to provide accounts of data. Some commentators have taken issue with this claim. SMYTHE & KOLERS seem to have mistakenly assumed that because we are unsure about the status of some of the aspects of our model, we are unsure about all of the aspects and hence cannot use the model. In fact, only a few properties of the model fall in the class of being "theory neutral" (See Hesse 1963), and these properties serve a useful role in promoting further research. HAYES-ROTH argues that the value of the model is dependent on how well we understand perception, image generation, and image representation, and that we have failed to provide adequate accounts for any of these capabilities. Not only does Hayes-Roth not elaborate this out-of-hand dismissal, but he seems to contradict an argument he made against Anderson (Hayes-Roth 1979) in which he terms a "fallacy" the notion that "incompletely operationalized theories are neither testable nor scientifically important." In addition, Hayes-Roth has misunderstood how our model accounts for particular data. As he notes, we argue that image inspection is facilitated by larger images, while image rotation is hindered by larger images. According to Hayes-Roth, we have a contradiction in principles since both image inspection and rotation are processes, and therefore processes are both facilitated and hindered by larger images. This argument is obviously flawed, and is analogous to the following one: All good football players are over 200 pounds. All good swimmers are under 200 pounds. Therefore, all good athletes are both over 200 pounds and under 200 pounds. In the case of size effects on image processes, we never use size per se as an explanatory construct, but rather appeal to properties that the theory confers on larger images. Regarding inspection, smaller images are poorly resolved and hence require zooming in before a part can be seen or inserted; zooming in is not required with larger images, and hence they are inspected more quickly. With respect to rotation, larger images occupy a larger portion of the visual buffer. The processor works only over that part of the buffer that is occupied by the to-be-rotated image. In the model, more cells of the surface matrix are occluded by larger images, and hence more operations are necessary at every iteration when it is rotated – requiring progressively more time to rotate larger images greater amounts. There is no contradiction here.

The third dimension. Several commentators (ABELSON, HAYES-ROTH, HINTON, MORAN, NEISSER, RICHARDSON, SHANON, WALTZ) have questioned the two-dimensional, quasi-pictorial structure of the surface display medium (the surface matrix) in the simulation. Here is a perfect example of a case in which a certain aspect of a model is unclear vis-à-vis the theory, and becomes the subject of further empirical investigation. Is the surface matrix two dimensional for

reasons of convenience (because we have only examined images of two-dimensional patterns so far), or should we claim that the medium underlying images, even images of three-dimensional scenes, is in fact two dimensional? It may come as a surprise to some that, as a result of a series of experiments addressed to this issue (Pinker and Kosslyn 1978; Pinker and Finke, in press; Pinker 1979 in press), we propose that the structure underlying images *is* functionally two dimensional (or 2½ D; see Marr and Nishihara 1978), as it is modeled in the current simulation. The telltale signs of a 2D representation of a 3D scene are perspective effects: dilation and contraction of size with changes in distance, invisibility of occluded objects, distortions of shape with changes in orientation, and other effects that depend on the viewer's vantage point. And most of these perspective properties can be shown to be true of the patterns depicted in images (see Pinker and Finke, in press, for a review of the evidence). As a result, we did not convert the surface matrix into a three-dimensional structure, since that would necessitate simulating internal "light rays," and a "lens" and "retina" for the "mind's eye," in order to give rise to perspective effects. That would clearly be an implausible arrangement.

Nonetheless, it seems likely that the 3D structure of a scene is preserved in memory, and the results of the Shepard and Metzler and Pinker and Kosslyn experiments suggest that images of 3D objects can be transformed smoothly in three dimensions (by, respectively, rotation in depth or scanning in depth). To account for these facts, we have suggested that the third dimension is represented at the "deep" level of the simulation, and that transformations in 3D space are performed at that level. There are several ways in which this could be modeled (as is discussed at length by Pinker 1979, chapter 8), none of which presents the simulation effort with any insurmountable obstacles (HINTON, MORAN and WALTZ notwithstanding).

Another possible modification is suggested by WALTZ and HAYES-ROTH when they mention Marr and Nishihara's "2½ D representation," a depiction in which each point is labeled with a vector representing the depth and surface orientation with respect to the viewer. It would be simple to add the third "half-dimension" to the surface matrix, by filling the cells with (depth, tilt, slant) vectors instead of with dots (just as we could follow MORAN's suggestion by filling the cells with n-place – depth, tilt, slant, color, intensity, texture, . . . – vectors), if we wanted to simulate the representation of these properties in images. In fact, Pinker (in press) and Pinker and Finke (in press) proposed exactly that. However, we await further data before committing ourselves on this issue (see Pinker 1979, chapter 8, for arguments pro and con).

HINTON's theory resembles ours in that it proposes a two-level structure for images of 3D objects, one representing the object's intrinsic shape, the second representing the perspective appearance of the object from a given vantage point. Our chief disagreement with Hinton concerns the form of the perspective or viewer-specific representation, since in our account, as in his, the deeper representation of shape includes a propositional component. The disagreement centers upon whether the viewer-specific representation is more like a description in a set of propositions or a depiction in an array. Though Hinton's available presentation is necessarily sketchy, it seems to us that the evidence favors our model over his (as we understand it). Pinker (in press) showed that people can scan an image in such a way that they take proportionally longer to scan between objects separated by greater distances in the 2D projection of the display. It is not obvious how this could work in Hinton's propositional scheme, in which the relation of each point relative to the viewer (but not the metric relations among all parts) is represented. Similarly, Hinton must explain the other findings concerning our proposed characterization of the surface image: the maximum visual angle of an image, the effects of image size on ease of part detection and on memory, and so on.

Three findings, in particular, seem very hard to reconcile with HINTON's theory. First, Keenan and Moore (1979) have shown that people have difficulty remembering objects that were concealed in their images, relative to their memory for unconcealed objects. This contradicts Hinton's suggestion that there is no reason to believe that "we perform any of the hidden-line-removal computations that would be necessary for generating a 2-D array." Second, if an object's intrinsic shape is encoded in a perspective-independent format at the deeper level, why do people mentally rotate one observed object into correspondence with another before determining whether or not they have the same shape (Shepard and Metzler 1971)? To use Hinton's words, why doesn't the homunculus peek at the intrinsic shape descriptions behind the scenes, allowing him to come to an immediate decision? In our model, if objects' shapes are encoded as lists of (r,Θ, Φ) coordinate triples relative to viewer-specific axes, no direct shape match at the deep level will be possible. It will be necessary to depict the rotation of one of the objects in the surface matrix, and only then can the template-matching processes compute a match or mismatch of the shapes.

But the most extreme empirical counterexample to HINTON's model concerns the effects of size on image transformation rate. He claims that "mental transformations involve changing some of the parameters of the intrinsic and projective relations, while leaving the rest of the description intact. The parameters that are changed are just those that would change during perception of a changing scene." Consider now what happens when one rotates a large image instead of a small one. Surely the size parameter in a description of an object would not change as the object is seen to rotate in the picture plane, and thus Hinton would lead us to expect that the size parameter would be left undisturbed and would have no effect on the rate of *mental* rotation. Schwartz (1979) has clearly disconfirmed this prediction: larger images are rotated at a slower rate than smaller images, complexity held constant. This is exactly what one would expect if the rotation operation was performed on a set of dots in an array: configurations of dots covering greater areas would entail lengthier processing at each iteration.

Scanning. WALTZ asks whether scanning an image consists of a translation of the imaged pattern across the visual buffer, or a shifting of the region of the buffer attended to by the interpretive process. Kosslyn, Ball, and Reiser (1978) and Kosslyn (1978a) argued for the former process because they found that subjects could scan smoothly from an object in the center of an image to an object that was initially "out of view of the mind's eye," with no kink in the line relating scan time to the distance between the source and the destination of scanning. Similarly, they noted that it seems easy to scan mentally around the four walls of an imaged room, never "bumping into an edge." Unfortunately for scientific parsimony, it now seems necessary to posit the second type of scanning (moving the locus of attention across the visual buffer) as well. Pinker (unpublished) has shown that people can scan in three-dimensional space between perceived objects in a visible display without moving their eyes. Since it is unlikely that this was accomplished by moving a "ghost" image of the scene relative to a fixation point, the finding argues that the locus of attention of the "mind's eye" is what moved. And because our theory states that images and percepts are activated in the same structure, it seems hard to deny that the locus of attention can shift across a stationary image. This in fact corresponds to the intuitions of many subjects in image-scanning tasks, and is in no way incompatible with the first mechanism we posited, which shifts patterns across the visual buffer (see Pinker in press). Furthermore, studies by Shulman, Remington, and McLean (1979) and by Pinker and Dintzis (unpublished) suggest that attention can indeed be shifted smoothly across the visual field, as required if the second putative mechanism exists.

Topics. Many interesting questions about imagery have been raised in the commentaries, such as those of ABELSON, ANTROBUS, COOPER, HINTON, HUNTER, JOHNSON-LAIRD, LUCE, PAIVIO, PYLYSHYN, RICHARDSON, SHEEHAN, and WALTZ. Sheehan, for example, points out that a person's motivations and needs will affect processing. Abelson, Cooper, and Richardson have wondered about imagery in different subject populations; many authors worried about the actual use and role of imagery; and Johnson-Laird raised the question of what subjects do when they are not using imagery. We have two responses

to these queries: First, we must start somewhere, Rome wasn't built in a day, and even a journal of a thousand volumes must begin with a single page. Many of the questions raised problems we plan to consider eventually. We have given priority to the topic of how imagery represents information in memory because we feel that this is fundamental to the other topics. For example, we feel that it is most profitable to study individual differences within a nomothetic, general theoretical framework. Hence, our efforts to study imagery in different populations awaited the development of the present theory of "modal processing" (see Kosslyn and Jolicoeur, in press). Similarly, we are now also studying when people spontaneously use imagery; this has become a tractable empirical question because we have a theory that helps us predict the tasks in which imagery should be used spontaneously, and we have a host of empirical techniques for detecting imagery use when it occurs. We are not trying to discover when imagery use is logically necessary, however, as Richardson wants to know, but rather we are studying when imagery is in fact used. Kosslyn, Murphy, Bemesderfer, and Feinstein (1977) found that the relative speed of imaginal and nonimaginal processing dictates whether or not imagery will in fact be used in a task, even if the use of imagery is not logically necessary (see Kosslyn, in press, for an extended treatment).

Our second response is to JOHNSON-LAIRD's complaint that we have not studied nonimaginal processing. We have chosen to study the domain of mental imagery; to fault us for not studying something else seems out of court. One reading of this complaint, however, is that we have not studied crucial parts of the imagery-processing system. Another reading might be that we have defined the domain incorrectly. But one must decompose the mind into some kind of constituents; trying to study the whole man, as NEISSER seems to advocate, is a little like trying to get one's mouth around a giant apple: There is no way to dig one's teeth in, and one runs the risk of dying from lack of nourishment. While there is no guarantee that we have chosen a natural parse, a genuine constituent of the mind, only time and continued experimentation and theorizing will allow us to answer this question.

III. Metatheory

The following are responses to comments, claims, and objections that address the assumptions underlying our attempt to construct and test a theory of the mental imagery representation system.

The homunculus. Our theory posits that imagery consists of patterns of activation in an internal medium that functions as an array. Are we then guilty, as FELDMAN, HEIL, and KEENAN & OLSON suggest, of populating the skull with an infinite series of graduated homunculi, each one watching a TV screen inside the head of the next largest (cf. Seuss 1957)? No. Once and for all, the "homunculus problem" is simply not a problem. We thought this would be obvious, given that the theory is realized in a computer program, but it seems necessary to address this complaint again (see Kosslyn and Pomerantz 1977).

Our theory describes both internal structures *and* the processes that generate, interpret, and transform the data structures. These processes can be identified either with fixed, innate neural circuits, or with combinations of such circuits, which access information stored in some as-yet-undiscovered fashion in other collections of neurons. This is in direct analogy to computer organization, in which data structures in core memory are interpreted by specific processes, either hard-wired instructions or subroutines composed of a series of such instructions. As computer programmers have known for years, it makes perfect sense to talk about a routine "constructing," "looking at," or "altering" a data structure, and notions of question begging, circularity, or infinite regress simply do not arise. The reason is clear: The routines are not miniature replicas of the entire computer or computer program, each one duplicating the whole. Rather, each subroutine has a special function; each is "stupider," to use Dennett's (1979) term, than the whole computer. In turn, a subroutine may be composed of subsubroutines that are even stupider (i.e. perform even simpler and more specific functions), and so on until the "stupidest" of the processes can be identified with a piece of computer hardware (e.g. an "and" gate). Similarly, in the case of humans, it is assumed that the information processes we posit can be successively decomposed into simpler and simpler processes, the simplest of which will be identifiable with specific neural events. As Dennett put it, "one discharges fancy homunculi from one's scheme by organizing armies of idiots to do the work."

Internal representation of the world. HEIL recounts a familiar objection to picture-based theories of imagery (see Ryle 1949). The objection has two parts: a logical argument and an empirical claim. The argument is, roughly, that imagining an object is unlike perceiving that object because only in the latter case does the person stand in some relation to a second entity (i.e. the actual object perceived). That is, imagining an object is in fact not at all like *seeing* an object, but is more like the result of having *recognized* an object, that is, *knowing* what that object is. Perceiving different objects consists of doing the same thing to different objects; imagining different objects, in contrast, consists of doing different things, with no objects involved. Confusion between the two is what led us to model images as entities – namely as spatial patterns in an internal array. The empirical claim intended to buttress this argument is that images cannot be ambiguous; unlike the case of perception, we always know precisely what we are imagining.

We believe that the position that HEIL reiterates is seriously in error. It betrays on the one hand a lack of understanding of the task of scientific psychology, and on the other, a totally unfounded empirical claim about the properties of images. Heil's logical argument rests on a certain presupposition: The claim is that no internal representations as such occur during either perception or imagery; NEISSER and SMYTHE & KOLERS also make this claim, which rejects the standard paradigm of cognitive psychology whereby the mind is analyzed in terms of structure-process pairs. The major argument against positing internal representations is discussed in another section and found wanting. If perception occurs by processing representations arising from stimulation of the sense organs, then there is no qualitative distinction between perception and imagery, the latter consisting of processing of similar representations arising from long-term memory. Both perception and imagery involve doing the same kinds of things to different "mental objects." As the very existence of our simulation attests, there is nothing logically inconsistent in the idea that images are data structures that are processed in various ways. Heil's presupposition has, in turn, led him to blur two different senses of the word "recognize." Consider the following situation: Barry and Mark are identical twins. I encounter Barry on the street and mistakenly say "Hello, Mark." In one sense, I have recognized Barry's face, in that I have successfully assigned a pattern of retinal stimulation to an equivalence class corresponding to the correct "concept" (i.e. I do not say "Hello, John"). In another sense, I have not recognized Barry since it is not he, but Mark, who is in fact standing on the pavement. Clearly, students of "pattern recognition" are concerned with "recognition" in the first, but not the second sense (assuming that clairvoyance and telepathy are not factors here). Recognition in the second sense (the one that Heil speaks of) is simply not part of the scientific problem called perceptual "recognition." The same problem of characterizing internal processing would occur if no actual person were present, but only a hologram (see also Fodor, in press).

As to HEIL's empirical claim (frequently cited by other philosophers) that an image cannot be ambiguous (in the sense of a Necker cube, wife/mother-in-law, or duck/rabbit), this is simply egregious armchair psychology. First of all, let us grant (only for the purpose of the argument) that people never report that they do not "know" what they are imaging. This observation would have *no* bearing on the structure of images per se. It is entirely possible that images themselves might be ambiguous (e.g. in our model, the

surface matrix might be filled with dots depicting a duck/rabbit), but that other cognitive faculties (e.g. propositions or subvocalizations to the effect "I am seeing a duck") are activated in such a way that the subject is aware of only one reading of the ambiguous pattern. In fact, as every instructor of introductory psychology can attest, this frequently occurs in perception: observers often have difficulty seeing the different interpretations of an ambiguous figure.

In any case, the argument is beside the point. HEIL is mistaken: Images *can* be demonstrably ambiguous. Consider the following examples. Kosslyn (in press) describes an experiment in which different subjects were given different descriptions of an ambiguous pattern and were asked to image the pattern (e.g. two overlapping rectangles versus four squares abutting a central square). All subjects correctly answered questions about the presence of parts in the imaged pattern (e.g. a rectangle), even though different parts were derived from different ways of parsing the figure. (However, subjects did require more time to "see" a pattern if it was not defined by the parse given in the initial description.) Thus, images *can* be ambiguous, and people can reinterpret them if need be. Contrary results reported in the literature (e.g. the Reed studies cited by HINTON may be traced to the excessive complexity of the imaged figures (see our discussion of Hinton's commentary for an explanation of such complexity effects). In addition, Kosslyn and Alper (1977) had subjects image pairs of objects, one of which was to be imaged so small that it appeared as a speck in the image. Such patterns can be considered ambiguous in that a rabbit with a speck on its back could be interpreted as a rabbit supporting a minuscule typewriter, a rabbit supporting a minuscule Volkswagen, a rabbit supporting a minuscule breadbox, and so on. And in fact, subjects had difficulty in remembering which object it was that they had imaged at the size level of a speck, empirically supporting the notion that their images were ambiguous in the way we described (see Kosslyn and Alper 1977 for details of various experimental controls supporting this conclusion, and Keenan and Moore 1979 for elegant supporting evidence). Finally, it has long been realized that any two-dimensional pattern is inherently ambiguous as a depiction of a three-dimensional scene, since many 3D scenes could have given rise to the same 2D projection. Pinker and Finke (in press) have shown that when subjects image a display of objects suspended in three-dimensional space, they can "see" both the three-dimensional structure of the display *and* the two-dimensional geometric shape inherent in the frontal projection of the display (see also Pinker in press). This corresponds to the well-known "railroad tracks" ambiguity in perception: I can see that the tracks in front of me converge toward the horizon, but I can also see the same tracks as parallel at every distance.

Finally, several commentators – e.g. HAYES-ROTH and JOHNSON-LAIRD – attribute to us the claim that human images invariably preserve the Euclidean spatial characteristics of the things in the outside world of which they are representations. This is not our view. As of now, we are uncertain about how much spatial distortion the imagery system permits or introduces in uncontrolled situations. We are uncertain about this because we have given little attention to the general nature of the relation between the geometric properties of mental images and the geometric properties of the things they represent. Instead, we first attempted to show that images *can* represent Euclidean properties of spatial displays. But for all we know, a New Yorker's mental image of the United States might correspond to Saul Steinberg's famous map in which the distance from Manhattan's Ninth Avenue to the Hudson appears equal to the distance from Nebraska to the Pacific. Thus, our subsequent experiments have been designed to control this degree of freedom. That is, we have generally tried to restrict the mental images used in our experiments to those that do preserve Euclidean spatial relations if such relations can be preserved at all. Our claim is that the "image on the screen" is processed as a Euclidean spatial representation, regardless of what spatial features it usually has in common with what it is supposed to represent in the world. Perhaps the confusion here stems from an ambiguity in the term "representation": On the one hand, it is taken as an internal data structure embodying information that need bear no determinate relation to the external world, and on the other hand, it is taken as information that is *true of* the external world. In the first case, the focus is on the nature of the representation as an entity in its own right, whereas in the second case the focus is on the mapping between the representation and the referent. When we say that images are spatial representations, we are using the term in the first of these two senses.

Images as epiphenomena. We labeled the first choice point in the decision tree on the nature of images "epiphenomenal versus not epiphenomenal." On the basis of experimental data, we chose the latter option, and we went on to model the image representation as an array. PYLYSHYN, MORAN, and KOHLERS & SMYTHE, among others, correctly point out that the epiphenomenal/nonepiphenomenal issue is logically distinct from the array-versus-propositions issue. Our conflating the two was intentional. Though we are certainly in no position to explain the origin or nature of the conscious experience of images, we subscribe to the view that these experiences must somehow be related to functional cognitive states, that is, to the activity of processes acting on data structures. Therefore, if people experience certain things, we assume in general that this is because of certain properties of their internal representations and processes. When people introspect, they report that images seem distinct from other forms of thought, that distance, size, and orientation are implicitly represented whenever objects are imaged, that images have a bounded size and grain, and so on. Of course, these experiences *could* have arisen from many different sorts of representations, but an array is the most parsimonious one.

Now, the question becomes: Are the particular representations underlying the conscious experience of imagery actually *used* in cognitive processing, or are they activated as incidental concomitants of the processing of *different* representations? The data that we and others have collected suggest that these representations *are* in fact used: When people solve tasks involving the conscious use of imagery, they take more time to "see" small parts of imaged objects, proportionately more time to scan greater imaged distances, and so on. These response time data, like the properties apparent to introspection during imaging, are most perspicuously explained by positing that imagery consists of the processing of patterns in an arraylike medium. Therefore, we claim that the representations underlying the experience of imagery are *not* epiphenomenal, and furthermore, that the aspects of those representations that account for properties of the conscious experience of imagery (e.g. the size and grain of the medium) can also be used to explain the information processing that underlies the task performance in these experiments. (Of course, we do not know to what extent functional cognitive states *can* explain conscious experience in the long run, but they seem to be the best explanations anyone has at the moment.) Note that our conclusion that the experience of imagery indexes properties of the underlying functional states is an empirical *discovery;* we do *not* subscribe to the view that introspections in general are a royal road to characterizing the format of internal representations.

Incidentally, KEENAN & OLSON claim that our model does not make images nonepiphenomenal since the CRT could be unplugged and no changes in processing would ensue. But this is not true. We have used the expression "CRT metaphor" as shorthand for the protomodel that guided the initial research phase. The metaphor was intended to include not just a CRT screen but also a processor that works over the display and interprets the patterns depicted on it. One way to think about this in mechanistic terms is to posit a TV camera aimed at the screen and connected to a pattern-recognition microprocessor (that, say, registered the presence of particular configurations of lines and angles – note that this camera would *not* lead to another screen which in turn needed another camera and so on, as discussed in our section on the homunculus). Of course, the actual phosphor-and-glass, light rays, and TV camera are superfluous to the analogy, and we could easily scrap them and let material in an array serve as the input to the pattern-recognition program. And this

hypothetical arrangement in fact evolved into the computer simulation model that replaced the CRT metaphor in our thinking.

Perception. Imagery and perception have long been considered to be cut from common cloth, and thus it is not surprising that many of the commentators on our target article had something to say about the relation between the two. FELDMAN, HAYES-ROTH, and HINTON seem to think that one cannot really study imagery without studying perception, and SHEEHAN argues that perceptual controls are desirable for drawing inferences about structure; ANTROBUS and HANNAY question whether studying just *visual* perception is enough; ABELSON and DE VEGA wonder about the purported connection between visual imagery and visual perception; LUCE, MORAN, and NEISSER wonder whether the "visual buffer" posited by our theory makes sense if it is supposed to be shared with perception; several of our critics claimed that imagery occurs after the stage of pattern recognition; and Luce and WALTZ wonder why we don't suppose that "image transformations" are used in perception as well as in imagery.

We have tried to study how images serve as data structures in memory that may be processed in various ways. Thus, we have explored the nature of these data structures in their own right. The question of how many components of the imagery system are also used in perception is a related, but separate, question. Although showing that imagery and perception are alike in some regard is often interesting in and of itself (for example, see Finke and Schmidt 1977; Shepard and Judd 1976), and serves to implicate the same system in both kinds of processing, it does not *necessarily* help one to specify the nature of the mechanisms involved in either system. For example, similar data might be obtained in imagery and perceptual tasks because in both cases propositions (or arraylike depictive representations) are processed. Thus, simply running perceptual controls will not necessarily tell us anything about image structure. Without having some prior idea of what will prove relevant or important, we may not gain much understanding of imagery simply by studying perception. Nor can we easily apply constraints from perception to an imagery theory.

It seems likely, in any case, that imagery *will* in fact share common structures and processes with perception in the same modality. Those who deny this have to explain the fact that forming a visual image disrupts visual perception more than auditory perception, but the reverse is true when an auditory image is formed (see Segal and Fusella 1970). In addition, recent work by Finke (1979) and Finke and Schmidt (1977) demonstrates that perceptual phenomena such as visual-motor adaptation and orientation-contingent color aftereffects (the McCollough effect) can be produced by having imaged stimuli substitute for perceived ones. Thus, it will not be surprising if the reason many of the components of a theory of imagery have their particular properties is that those components are also used during perception and are tailored to its demands. But perception will involve many components that are not shared by imagery, and vice versa, so merely studying perception will not necessarily further an understanding of imagery per se.

The notion of a visual buffer as we have characterized it is not implausible as part of a theory of perception. In fact, the so-called 2½-D sketch posited in the Marr and Nishihara (1978) theory of visual processing is functionally similar to a surface image as we have characterized it, in that this representation preserves topographic information about the appearance of surfaces of objects. The visual buffer we have posited is a structure that has certain functional properties (outlined in detail in Kosslyn, in press). Whether those properties are derived by aggregating over collections of feature and location detectors or by some complex processing of a Fourier transform is not to the point: At the functional level, an arraylike "visual buffer" provides the most perspicuous account of a large set of data, in perception as well as imagery (see Finke and Kosslyn, in press; Kosslyn 1978-a). It will be important to discover exactly how the properties of this cognitive structure are related to the underlying physiology, but whatever this relation may turn out to be, on a higher level of analysis the brain behaves as if it has a representational structure with the properties of the visual buffer.

Several commentators, notably DE VEGA, HINTON, MORAN, NEISSER, and WALTZ, asserted that images must occur at some stage *following* pattern recognition, not prior to it, as we have claimed. Let us clarify our claim. Consider the oversimple but not ridiculous view that vision is supported by a sequence of data structures starting with the pattern of receptor activation and culminating in a set of semantic or conceptual propositions describing what is seen. On this view, "pattern recognition processes" transform the more peripheral representations into increasingly central ones. The imagery debate (to oversimplify further) can be summarized as a disagreement over which of the data structures in the sequence support mental images (and consequently, which particular set of pattern recognition processes inspect images). At one extreme is the view (held by no one we can think of, and certainly not us) that images occur in the retina, and that processes such as contour enhancement and color normalization occur in imagery as well as in perception. At the other extreme is the propositionalist position that images are purely symbolic structures no different from those underlying abstract thought, and that *no* visual pattern recognition processes apply. Our position is that images are representations like those that occur in intermediate stages of visual processing, and that *some* visual pattern recognition processes can operate over them (e.g. detecting geometric shapes or parts of animals that do not receive explicit propositional encoding in long term memory). For example, the contents of the visual buffer may already be parsed into "Gestalt wholes" and interpreted in a perceptual sense, but not yet labeled or identified with semantic categories. (For evidence supporting this view, see our above discussion of "ambiguous images.") The same is true of the underlying "literal" representations, which is why we use the word literal in quotes (in answer to KEENAN & OLSON's question): the stored information is not raw and uninterpreted but is not simply "symbolic" either.

Why did we shy away from ascribing all of the functional capacities of imagery to perception? The answer is simple: we were disinclined to accuse our fellow human beings of being solipsists. The image transformations, for example, alter the representations underlying our phenomenal experience – which does not seem to happen in normal perception, and in fact would characterize a rather extreme hallucination. We found LUCE's and WALTZ's intuitions rather congenial, however, and are willing to entertain the hypothesis that some of the image generation and transformation processes are in fact used to normalize patterns in perception. However, there remains the problem of distinguishing between the different ways in which these processes affect our *images* (e.g. by using images as templates to be matched against percepts – see Kosslyn, in press), as opposed to the way they affect our *percepts* per se (if they do at all).

Images and propositions. Several commentators asserted directly or indirectly that images are best characterized as something akin to sets of symbolic propositions, and not as patterns in an array (HEIL, HINTON, MORAN, SMYTHE & KOLERS). Some of the commentators argued that there is no real distinction between descriptive and depictive representations. As ANTROBUS pointed out, the locations of points in a matrix can be described propositionally. But to do so is to confuse what Putnam (1973) calls "the parent of an explanation" with the explanation itself. The locations of individual points fail to tell one about higher-order relations (deriving from the geometrical properties of the array) that convey information. A list of which cells are filled and unfilled on a CRT screen would not reveal what picture was being displayed. At the level of functional properties there are real differences between the different kinds of representations; for example, the symbols in an array (points at particular locations) are not arbitrarily related to the properties of the object or scene itself, whereas the symbols used in a proposition are arbitrarily related to the thing being represented (see Kosslyn, in press, for a detailed characterization of the differences between

propositional and quasi-pictorial formats at a functional level). Further, a depiction need not be first-order isomorphic, as KEENAN & OLSON claim: Rather, the representation must depict vis-a vis the processes that interpret it. In the computer, for example, the points in an array depict although there is no first-order isomorphism between the physical representation in core and the points on the depicted object(s).

Finally, none of these commentators, however, attempted to account for the data that, we argue, favor an arraylike representation. We cannot emphasize strongly enough that the study of imagery has progressed beyond the point where a priori arguments and intuitions can be decisive. There is a rich and growing set of empirical findings that must be addressed by anyone who takes a stand on imagery. Ex cathedra statements that imagery must be one way or another serve no purpose; bald assertions do not constitute arguments or explanations.

We agree that the functioning of the brain could be described exclusively in terms of a propositional representation system, but this does not mean that this is the *correct* way of describing it. The brain is a physical device that operates in particular ways and not in others, and we view our job as one of discovering how it actually functions. That is, there *is* a "fact of the matter": A functionally spatial medium, supporting representations that depict, either does or does not exist.

Demystification. PAIVIO has pointed out that one way in which something is demystified is by the discovery of factual information. This is certainly true, and Paivio deserves much of the credit for making the study of imagery respectable again. In fact, Paivio almost single-handedly showed that imagery could be studied with rigor and that scientific theories of imagery could be formulated. But there is a second way in which topics like imagery become demystified, and this is through conceptual clarification. In fact, several commentators wonder whether our notion of imagery is so unclear as to be a major source of confusion in our research. Thus JOHNSON-LAIRD wonders whether various assertions about imagery are testable hypotheses or parts of the definition of an image. (We are reminded of the longstanding controversy over whether $F=ma$ is a definition or a testable hypothesis in Newton's mechanics.) and HEIL argues that we have violated the maxim, "in investigating and theorizing about a phenomenon . . . it is advisable to begin with a reasonably clear idea of what it is one is investigating." The problem with a maxim like this is that if one has a reasonably clear idea of something like imagery, the need for elaborate experimentation and theorizing largely disappears. A theory of image processing will ultimately tell us what images in human cognition are, just as the theory of the electron tells us what electrons are. When research on electrons – more precisely on cathode rays – began, physicists had little idea of what electrons were. (Cathode rays were hypothesized to be streams of some sort of negatively charged particles, and a dispute, somewhat like the one about imagery, arose when physicists working close to Maxwell's theory countered that the rays were some sort of electromagnetic waves.) But this did not impede the ultimate development of the theory of the electron. One way in which one is forced to think more concretely, if not more clearly, about the nature of representation and processing is by attempting to formulate a detailed simulation model. The detailed simulation model is itself evidence that whatever confusion there is, it is not standing in the way of research. What must be shown to counter this evidence is that our model trades systematically on some ambiguity in the notion of imagery. One virtue of simulation models is that they tend to safeguard against such a possibility. Whether or not this exercise has its intended end is, of course, as yet an open question. As we have argued in the target article, the test will be whether this project continues to produce empirical results that accumulate to convey a coherent picture about the imagery system. We wish to note that "conceptual clarification" about imagery will most profitably be performed only in close conjunction with data collection and explanation; if the history of imagery has taught us anything, it is that armchair theorizing is not likely to lead to much progress.

Ecological validity. All of our data were obtained in very "artificial" situations, as NEISSER and SHEEHAN pointed out. These situations do not resemble those in which imagery presumably occurs during the course of everyday life. This kind of approach is not unusual in science, however; since most "natural" phenomena are overdetermined, one must go into the laboratory and create idealizations in which the number of active variables is sharply reduced. In so doing, of course, one runs the risk of having eliminated the important aspects of a phenomenon in the interests of tractability. But there seems no way around this risk; to try to study the mind as a single working entity in its natural surroundings is not feasible. Of course, NEISSER might not subscribe to this extreme view and may instead merely stress the importance of studying variables that are related to those encountered by the organism in its natural habitat. If one's results can in fact generalize back to the original phenomena of interest and to other natural phenomena, they seem to fulfill this requirement. HUNTER and WALTZ suggest that the present approach at least has some promise of accomplishing exactly this kind of generalizability.

Theories and models. Some of the issues raised in the commentaries seem to us to turn on too literal an interpretation of the model. For example, LUCE, MORAN, and HAYES-ROTH worry about the kind of coordinates used in the long-term memory (LTM) files. In our simulation model, we represent "literal" information as a list of r, θ coordinates. However, we did not mean to propose that people store images using exactly this representation. Our only theoretical claim regarding this LTM data structure is that it is sufficient to generate short-term memory (STM) images without loss of metric information present in the original physical objects(s) or scene(s). Both Cartesian and polar coordinate lists fit this description. We chose a polar coordinate representation because we found that people could generate images at arbitrary size levels, and a polar coordinate system makes this easier. But the actual coordinate structure of the LTM representations – and the visual buffer itself, for that matter – remains an open question at the present time.

In general, criticisms that focus on some comparatively concrete detail of our model do not cut as deeply as other lines of criticism. As SHEEHAN notes, a central point of the model is to characterize results pertaining to "image processing" in a unified way that promotes further discoveries. Criticisms that focus on a detail of our model have the most bite when they provide reasons for thinking either that the research advantages of the analogy have been largely exhausted or that the analogy has become systematically misleading. The working criteria we use to evaluate the model – that is, the criteria discussed at the end of the target article – are intended to be responsive to these two possibilities.

SHEEHAN considers it "especially critical" that we distinguish between the substantive theoretical claims and the mere metaphors in the model – that is, between the positive and negative analogy. Clearly, we do not want to claim that an inability to distinguish between the positive and negative analogy is an ideal situation. But generating questions about whether a given feature belongs to the positive or negative analogy is one way a model helps guide a research program. Experimental research centered on the model can be viewed in part as an endeavor to explore and sort out the as yet undetermined part of the analogy – what Hesse (1963) has called the "neutral analogy." However, the model is not a static entity which at any given time is fully endowed with all of the properties it could have. Rather, the model is continually being developed, by being both further refined *and* expanded. In fact, our main strategy has been to consider alternative ways of developing the model and then conducting experiments to discriminate among them. The result is a program of experimental research that is closely tied to the model – initially to the protomodel and subsequently to the simulation model. COOPER objects to this narrow focus. She suggests that the focus on

model-derived questions may prevent other significant issues pertaining to imagery from being addressed. We agree with her that the strategy does this, but we differ from her in that we consider this its principal virtue. The difficulty with significant issues is that they are unlikely to give way to experimental investigation unless the set of moves one might make in response to them is properly constrained. A standard explanation as to why experimental research is more productive in highly theoretical sciences is that theoretical considerations sharply limit both the range of possible answers to questions and the range of possible interpretations of experimental results. What theory supplies in the more mature sciences, the model is supposed to supply in the current research. If the model is a good one, it will keep research focused on issues that will yield to experimental investigation at the time they are considered. As such, the model ideally should dictate the sequence in which issues are addressed. It should prompt a question only when we are in a good position to address it, and it should postpone consideration of other questions, however interesting they may be prima facie.

Some qualifications are obviously needed at this point. We agree that the issues COOPER lists as significant must ultimately be addressed. The only point on which we disagree is when. Also, we concede the risk that our approach will yield purely artifactual lines of research. This is why it is important to recognize when the model has ceased to be fruitful. One sign of this is when new experimental results cannot be incorporated into the model without backtracking (and hence the model fails in its coherence and/or generality, to use FELDMAN's terms). When this happens, the model is prompting questions that in fact are not properly constrained. Our working criteria for evaluating the model are responsive to just such signs. But so far these criteria have given us no occasion to abandon the model.

SCHANK sympathizes with our view of our model. He goes on to remark that the model is an adequate theory so long as it produces the correct input-output behavior for the "right" reasons. We agree with this remark as far as it goes, but we may well not agree with Schank about what the right reasons are. He requires the theoretical framework to accomodate all data pertaining to input-output behavior. We agree with BRIDGEMAN and FELDMAN in additionally requiring the theory to accomodate data pertaining to neurophysiological realization. We are prepared to concede HAYES-ROTH's point that a simulation model based on a CRT metaphor and executed on a digital computer is likely to differ from what happens in the brain with respect to some aspects of mechanism. That is, we are prepared to concede that our simulation model will invariably retain some elements of negative analogy. In those respects in which it does, it will remain just a model regardless of how well it predicts input-output behavior. And in those respects it will ultimately become an impediment to research. Instead of describing behaviors in terms of their true generation, it will describe them *as if* they were generated in the manner the model says. A theory, unlike a model, would eliminate the *as if* element from the account. This is why we think the long-term goal is a theory of image processing and why we view the simulation model as just a substitute for theory during the early stages of research.

We agree with FELDMAN that, other things being equal, models that "have a demonstrable reduction to physical reality" should be preferred to those that do not. And in fact, an important constraint on a cognitive theory is that the different functional components be in some sense neurophysiologically distinct. For a description of how imagery processing works to be correct, a function-structure isomorphism is required. However, the sort of neurophysiological data needed to establish this are not available. Compare the imagery-processing problem with that of the language processor. A century of neurophysiological and behavioral research on aphasia has, at least until recently, contributed little to the effort to describe how language is processed [cf. Arbib & Caplan: *BBS* 2(3) 1979]. In the case of imagery, far less data on brain-damage-induced pathologies are available. As such data become available, they clearly should be reviewed in the light of the model, and vice versa. But as things stand, we have no reason to think that placing a heavier emphasis on questions of neurophysiological realization will help us now to gain further insights into what is going on in imagery processing.

Finally, PYLYSHYN's claim that the "explanatory power" of our theory continues to flow out of the physical CRT model is an odd one. He says, "it is only when you have a *real* distance that the time taken to traverse it *must* vary with the magnitude of the distance." But a similar criticism could be applied to *any* scientific theory that relied on a physical model at some point in its development – the theory could *never* be as "explanatory," in Pylyshyn's sense, as the original physical model. It would make no sense to impugn a graph-searching theory of long-term memory by pointing out that "it is only when you have a *real* tree that the time taken to get from one branch to another *must* vary with the number of intervening branches." Similarly, it would be absurd to say of Bohr's atom that "it is only when you have a *real* solar system that there must be a massive, solid body surrounded by smaller orbiting satelites."

The debate. PYLYSHYN accuses us of "completely misunderstanding" the nature of the opposition to our position. To that charge, we plead "not guilty." There is an ambiguity in Pylyshyn's accusation, however. If he has in mind models based on the picture metaphor when he refers to "models such as [ours]," then our characterization of the opposition seems entirely accurate. The reader has only to look through the commentaries to see that we can hardly be said to have slandered our critics in representing them as maintaining that imagery is not a well-formed domain, that the concept is incoherent or vague, and so on. Moreover, one need go no further than the title of Pylyshyn's 1973 paper, "What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery," to realize that Pylyshyn himself has argued that the construct was itself paradoxical and flawed (for specific quotes on imagery being epiphenomenal, see page 6; on the notion of imagery being vague and incoherent, page 2; on imagery lacking heuristic value, page 8; on imagery not being a well-defined domain in its own right, page 21). Kosslyn and Pomerantz (1977) summarize Pylyshyn's arguments and show that none of them is compelling.

On the other hand, PYLYSHYN might concede that our model answers many of these old criticisms of the picture metaphor, and that we have misunderstood *new* arguments, centering around the notion of "cognitive penetrability," that have been directed specifically against our theory. We responded to these arguments in the target article, but for the most part Pylyshyn has declined to discuss our rebuttal and to show how we allegedly "misunderstand" the criticisms. For example, he ignores our argument against identifying explanatory accounts with accounts stated at the level of "impenetrable" processes; he ignores our discussion of *which* components of the imagery system we hold to interface with knowledge and belief systems and thus to allow penetration; and he ignores our arguments about the proper evaluation metric for models in sciences at an early stage of development.

Let us set the record straight, then, and outline our understanding of, and responses to, PYLYSHYN's "penetrability argument." Pylyshyn seems to use it in three different ways, switching back and forth without warning, and we think it would be instructive to spell these out.

1. The strongest position: Accounts involving cognitively penetrable processes are just wrong. PYLYSHYN implies that he is making this argument in his choice of a title for his commentary, but questions of relative "right" and "wrong" (i.e. failing to account for certain data as well as an alternative theory does) are noticeably absent from the commentary. The strong argument entails that we *know* that higher-order functional units of the mind have no access to or influence over the internal workings of lower-order units. We are utterly baffled as to how anyone could insist that this principle is true, given our current knowledge of how the mind is organized. Certainly, it has no basis in neurophysiological considerations (as BRIDGEMAN points out), nor do artificial intelligence considerations force it on us (see, e.g. Minksy 1979), and even the most persuasive a priori argument for a hierarchical organization of the mind (Simon 1969) stops short of

claiming that it must be strictly or totally decomposible into hermetically sealed units. (See also Luce and Green 1972; Green and Luce 1973 for arguments that cognitive factors can penetrate even to the level at which neural pulse trains are "interpreted.")

2. A weaker position: Accounts that involve cognitively penetrable processes are not sufficiently explanatory. That is, since the executive is extremely flexible (on all accounts), any theory that allows the executive to interfere in the operation of other cognitive processes will have so many degrees of freedom that constrained, strongly falsifiable explanations will be impossible. First, we see no reason why it may not be true that the executive *can* interfere with all the processes we call "cognitive," and if it is, there is little that we or anyone else can do about it. As George Miller put it, "The human being was not created for the benefit of psychologists." Note that we are fully aware of the dangers of subscribing to an extreme version of the hypothesis that the mind is a complex and undifferentiated whole, with no autonomous functional units: if so, an explanatory science of cognitive psychology would be impossible. But, to paraphrase Einstein, we do not feel that God has been *that* malicious, only subtle enough to have allowed the executive *some* powers of penetration into certain other components. If true, this moderate position might preclude the possibility of there being strongly explanatory accounts of certain cognitive phenomena, analogous to the accounts found in the physical sciences. However, it will allow some explanations to be more motivated and less ad hoc than others, especially if the *loci* and *nature* of the penetration are circumscribed in the theory – a task we began in the target article. For example, our commitment to the notion that the "visual buffer" supports representations involved in visual perception introduces a host of constraints and falsifiable predictions (see, e.g. Finke and Kosslyn, in press; Pinker, in press; Schwartz 1979) that are sufficient in themselves to belie Pylyshyn's claim that no explanatory power is lost in allowing the executive to mimic the privileged properties of the buffer.

3. The weakest position: Accounts involving cognitively penetrable processes raise methodological problems. This point is a version of RICHMAN ET AL.'s criticism, that some of the experimental methods we use may be flawed in certain ways. If this is the crux of PYLYSHYN's argument, then we do not understand why he felt compelled to write such a sweeping denunciation of the theory. For surely the question now becomes one of patching up the methods to eliminate the alleged alternative explanations, or to develop new methods that will serve as mutually agreed-upon tests of the relevant theoretical assertions. In the section in the target article and in this response, entitled "Demand characteristics," we have begun to do so, and it appears that our claims are or will be vindicated (see, e.g., Spoehr and Williams 1978; Shulman, Remington, and McLean 1979; and Kosslyn, in press, for experiments that support our conclusions about scanning without using explicit scanning instructions). Though debates about these methodological issues may lack the excitement of knock-down, drag-out battles over whether certain theories are truly "explanatory" (or, for that matter, "just plain wrong"), we feel that these pedestrian discussions are more fruitful at this stage.

REFERENCES

Abelson, R. P. (1968) Simulation of social behavior. In *Handbook of social psychology, vol. 2*, ed. G. Lindzey and E. Aronson. Reading, Mass.: Addison-Wesley. [RPA]

Anderson, J. R. (1976) *Language, memory and thought*. Hillsdale, N.J.: Erlbaum Associates. [PNJ, SMK, TPM, CLR]

(1978) Arguments concerning representations for mental imagery. *Psychological Review* 85:249–77. [PNJ, JMK, SMK, CLR, PWS]

(in press) Further arguments concerning mental imagery: a response to Hayes-Roth and Pylyshyn. *Psychological Review*. [SMK]

Anderson, J. R., and Bower, G. H. (1973) *Human associative memory*. New York: V. H. Winston & Sons. [SMK]

Arieti, S. (1955) *Interpretation of schizophrenia*. New York: Bruner-Mazel. [BS]

Asch, S. E. (1956) Studies of independence and conformity: a minority of one against a unanimous majority. *Psychological Monographs* 70 (whole issue). [SMK]

Ashton, R.; McFarland, K.; Walsh, F.; and White, K. (1978) Imagery ability and the identification of hands: a chronometric analysis. *Acta Psychologica* 42:253–62. [AR]

Attneave, F. (1968) Triangles as ambiguous figures. *American Journal of Psychology* 81:447–53. [GH]

Bahill, A. T., and Stark, L. (1979) The trajectories of saccadic eye movements. *Scientific American* 240:108–17. [SMK]

Baylor, G. W. (1971) *A treatise on the mind's eye*. Ph. D. dissertation, Carnegie-Mellon University. [SMK]

Beech, J. R. (1979) A chronometric study of the scanning of visual representations. Doctoral thesis, The New University of Ulster, Northern Ireland. [IMLH]

Beech, J. R., and Allport, D. A. (1978) Visualization of compound scenes. *Perception* 7:129–38. [SMK]

Boggess, L. C. (1978) Computational interpretation of English spatial prepositions. PhD. thesis, University of Illinois. [DLW]

Bower, G. H. (1970) Analysis of a mnemonic device. *American Scientist* 58:496–510. [DLW]

(1978) Representing knowledge development. In *Children's Thinking: What Develops?*, ed. R. S. Siegler. Hillsdale, N.J.: Erlbaum Associates. [SMK]

Bridgeman, B. (1971) Metacontrast and lateral inhibition. *Psychological Review* 78:528–39. [BB]

(1978) Distributed sensory coding applied to simulations of iconic storage and metacontrast *Bulletin of Mathematical Biology* 40:605–23. [BB]

Brooks, L. R. (1967) The suppression of visualization in reading. *Quarterly Journal of Experimental Psychology* 19:288–99. [PWS]

(1968) Spatial and verbal components of the act of recall. *Canadian Journal of Psychology* 22:349–68. [AR]

Bundesen, C., and Larsen, A. (1975) Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance* 1:214–20. [SMK]

Clark, H. H. (1969) Linguistic processes in deductive reasoning. *Psychological Review* 76:387–404. [JMK]

Conrad, C. (1972) Cognitive economy in semantic memory. *Journal of Experimental Psychology* 92:149–54. [SMK]

Cooper, L. A. (1975) Mental rotation of random two-dimensional shapes. *Cognitive Psychology* 7:20–43. [SMK]

Cooper, L. A., and Podgorny, P. (1976) Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance* 2:503–14. [SMK]

(1975) Mental transformations in the identification of left and right hands. *Journal of Experimental Psychology: Human Perception and Performance*. 1:48–56 [SMK]

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

Dennett, D.C. (1979) *Brainstorms*. Montgomery, Vt.: Bradford Books. [SMK, DLW]

DiVesta, F. J.; Ingersoll, G.; and Sunshine, P. (1971) A factor analysis of imagery tests. *Journal of Verbal Learning and Verbal Behavior* 10:471–79. [CLR]

Farley, A. M. (1974) VIPS: a visual imagery and perception system; the result of protocol analysis. Ph.D. dissertation, Carnegie-Mellon University. [SMK]

Finke, R. A. (1979) The functional equivalence of mental images and errors of movement. *Cognitive Psychology* 11:235–64. [SMK]

Finke, R. A. (in press) Levels of equivalence in imagery and perception. *Psychological Review*. [SMK]

Finke, R. A., and Kosslyn, S. M. (in press) Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance*. [SMK]

Finke, R. A., and Schmidt, M. J. (1977), Orientation-specific color aftereffects following imagination. *Journal of Experimental Psychology: Human Perception and Performance*, 3:599–606. [SMK]

Fodor, J. A. (1968) *Psychological explanation: an introduction to the philosophy of psychology*. New York: Random House. [SMK]

(in press) Methodological solipsism considered as a research strategy in cognitive psychology. *The Behavioral and Brain Sciences*. [SMK]

Fodor, J. D., Fodor, J. A.; and Garrett, M. F. (1975), The psychological unreality of semantic representations. *Linguistic Inquiry* 4:515–31. [PNJ]

Freud, S. (1958) *The interpretation of dreams*. London: Hogarth Press. [BS]

Friedman, A. (1978), Memorial comparisons without the mind's eye. *Journal of Verbal Learning and Verbal Behavior* 17:427–44. [UN]

Galton, F. (1883) *Inquiries into human faculty and its development*. London: Dent (1905). [AR]

Gibson, J. J. (1966), *The senses considered as perceptual systems*. Boston: Houghton Mifflin. [FH]

Goodman, N. (1968) *Languages of art*. Indianapolis: Bobbs-Merrill. [JH]

Green, D.M., and Luce, R. D. (1973) Speed-accuracy tradeoff in auditory detection. In *Attention and performance IV*, ed. S. Kornblum. New York: Academic Press. [SMK]

Haber, R. N. (1970) How we remember what we see. *Scientific American* 222:104–12. [DLW]

Hayes-Roth, F. (1977), Critique of Turvey's "contrasting orientations to the theory of visual information processing." *Psychological Review* 84:531–35. [FH]

(1979) Distinguishing theories of representation: a critique of Anderson's "Arguments concerning mental imagery." *Psychological Review* 86:376–820. [FH, SMK]

Hannay, A. (1971) *Mental images: A defence*. Atlantic Highlands, N.J.: Humanities. [AH]

Hebb, D. O. (1968) Concerning imagery. *Psychological Review* 75:466–77. [ZP]

Hesse, M. B. (1963) *Models and analogies in science*. London: Sheed and Ward. [SMK]

Hinton, G. E. (1979, in press) Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science*. [GH]

Hirst, W. (1976) Memory for proofs. Doctoral dissertation, Cornell University. [UN]

Hubel, D. H., and Wiesel, T. N. (1962), Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology* 160:106–54. [JSA]

Hunter, I. M. L. (1962) An exceptional talent for calculative thinking. *British Journal of Psychology* 53:243–58. [MLH]

(1977) Imagery, comprehension and mnemonics. *Journal of Mental Imagery* 1:65–72. [IMLH]

Huttenlocher, J. (1968) Constructing spatial images: a strategy in reasoning. *Psychological Review* 75:550–60. [JMK]

Johnson-Laird, P.N. (1979a) Formal semantics and the psychology of meaning. Paper presented at the Symposium on Formal Semantics and Natural Language, University of Texas at Austin. [PNJ]

(1979b) Mental models in cognitive science. Paper presented at the La Jolla Conference on Cognitive Science. [PNJ]

Jonides, J.; Kahn, R.; and Rozin, P. (1975) Imagery instructions improve memory in blind subjects. *Bulletin of the Psychonomic Society* 5:424–26. [UN]

Just, M.A., and Carpenter, P. A. (1976) Eye fixations and cognitive processes. *Cognitive Psychology*. 8:441–80. [TPM]

Kandinsky, W. (1947) *Point and line to plane*. New York: The Guggenheim Foundation. [BS]

Keenan, J. M. (1978) Psychological issues concerning implication: comments on "Psychology of pragmatic implication: Information processing between the lines" by Harris and Monaco. *Journal of Experimental Psychology: General* 107:23–27. [JMK]

Keenan, J. M., and Moore, R. E. (1979) Memory for images of concealed objects: a reexamination of Neisser and Kerr. *Journal of Experimental Psychology: Human Learning and Memory* 5:374–85. [JMK, SMK]

Kintsch, W. (1974) *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum Associates. [PNJ]

Kolers, P.A., and Smythe, W. E. (in press) Images, symbols, and skills. *Canadian Journal of Psychology*. [WES]

Kosslyn, S. M. (1973) Scanning visual images: Some structural implications. *Perception and Psychophysics* 14:90–94. [SMK]

(1974) Constructing visual images. Ph.D. dissertation, Stanford University. [SMK]

(1975) Information representation in visual images. *Cognitive Psychology*. 7:341–70. [SMK]

(1976a) Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval time. *Memory and Cognition* 4:291–97. [SMK]

(1976b) Using imagery to retrieve semantic information: a developmental study. *Child Development* 47:434–44. [SMK]

(1978a) Measuring the visual angle of the mind's eye. *Cognitive Psychology* 10:356–89. [SMK]

(1978b) Imagery and cognitive development: a teleological approach. In Children's thinking: what develops? R. S. Siegler, ed. Hillsdale, N.J.: Erlbaum Associates. [SMK]

(in press) *Image and mind*. Cambridge, Mass.: Harvard University Press. [SMK]

Kosslyn, S. M., and Alper, S. N. (1977) On the pictorial properties of visual images: effects of image size on memory for words. *Canadian Journal of Psychology* 31:32–40. [SMK]

Kosslyn, S. M.; Ball, T. M.; and Reiser, B. J. (1978) Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance* 4:47–60. [SMK, CLR]

Kosslyn, S. M., and Jolicoeur, P. (in press). A theory-based approach to the study of individual differences in mental imagery. In, *Aptitude, learning, and instruction: cognitive processes analyses*, ed. R. E. Snow, P-A. Fredrico, & W. E. Montague. Hillsdale, N.J.: Erlbaum Associates. [SMK]

Kosslyn, S. M.; Murphy, G. L.; Bemesderfer, M. E.; and Feinstein, K. J. (1977) Category and continuum in mental comparisons. *Journal of Experimental Psychology: General* 106:341–75. [SMK]

Kosslyn, S. M., and Pomerantz, J. R. (1977) Imagery, propositions, and the form of internal representations. *Cognitive Psychology* 9:52–76. [SMK]

Kosslyn, S. M.; Reiser, B. J.; and Farah, M. (submitted for publication) Generating visual images. [SMK, ZP]

Kosslyn, S. M., and Shwartz, S. P. (1977) A simulation of visual imagery. *Cognitive Science* 1:265–95. [GH, SMK]

(1978) Visual images as spatial representations in active memory. In *Computer vision systems*, ed. E. M. Riseman and A. R. Hanson. New York: Academic Press. [SMK]

Larsen, A., and Bundesen, C. (1978) Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance* 4:1–20. [SMK]

Lea, G. (1975) Chronometric analysis of the method of loci. *Journal of Experimental Psychology: Human Perception and Performance* 1:95–104. [SMK]

Luce, R. D., and Green, D. M. (1972), A neural timing theory for response times and the psychophysics of intensity. *Psychological Review* 79:14–57. [SMK]

Luria, A. R. (1968) *The mind of a mnemonist*, New York: Basic Books. [IMLH, DLW]

McCulloch, W. S., and Pitts, W. H. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5:115–33. [BB]

McKellar, P. (1963) Differences of mental imagery. *The Mensa Correspondence* 51:1–5. [AR]

MacLeod, C. M.; Hunt, E. B.; and Mathews, N. N. (1978) Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior* 17:493–508. [JMK]

Mamor, G. S., and Zaback, L. A. (1976), Mental rotation by the blind: does mental rotation depend on visual imagery? *Journal of Experimental Psychology: Human Perception and Performance* 29:263–91. [AR]

Marr, D. (1978) Representing visual information. In *Computer vision systems*, ed. A. R. Hanson and E. M. Riseman, New York: Academic Press. pp. 61–80. [DLW]

Marr D., and Nishihara, H. K. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society*, Series B 200:269–94. [FH, GH, SMK]

Marr, D., and Poggio, T. (1976) Cooperative computation of stereo disparity (memo 364). Cambridge, Mass.: MIT A.I. Lab. [FH]

Miller, G. A.; Galanter, E.; and Pribram, K. (1960) *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston. [PNJ]

Minsky, M. L. (1975) A framework for representing knowledge. In *The psychology of computer vision*, ed. P. H. Winston, New York: McGraw-Hill. pp. 211–77. [DLW]

(1977) Plain talk about neurodevelopmental epistemology. Paper presented at the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Mass. [BS]

(1979) *The Society of Mind*. Artificial Intelligence Laboratory, M.I.T. Unpublished manuscript. [SMK]

Mitchell, D. B., and Richman, C. L. (in press) Confirmed reservations: mental travel. *Journal of Experimental Psychology: Human Perception and Performance*. [CLR]

Moran, T. P. (1973) The symbolic nature of visual imagery. Paper presented at the Third International Joint Conference on Artificial Intelligence, Stanford University. [SMK]

(1973a) The symbolic imagery hypothesis: a production system model. Ph.D. dissertation, Carnegie-Mellon University. [TPM]

Morton, J. (1969) Interaction of information in word recognition. *Psychological Review* 76:165–78. [BS]

Neisser, U. (1976) *Cognition and reality*. San Francisco: W. H. Freeman. [JH, JMK, UN]

(1978) Anticipations, images, and introspection. *Cognition* 6:169–74. [JMK]

Newell, A., and Simon, H. A. (1972) *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall. [ZP]

Noy, P. (1973) Symbolism and mental representation. *The Annals of Psychoanalysis* 1:125–158. [BS]

Orne, M. T. (1962) On the social psychology of the psychology experiment: with special reference to demand characteristics and their implications. *American Psychologist* 17:776–83. [SMK, CLR]

Paivio, A. (1975) Perceptual comparisons through the mind's eye. *Memory and Cognition* 3:635–48. [SMK]

(1975a) Neomentalism. *Canadian Journal of Psychology* 29:263–91. [AR]

Palmer, S. E. (1978) Fundamental aspects of cognitive representation. In *Cognition and categorization.* ed. E. Rosch and B. B. Lloyd. Hillsdale, N.J.: Erlbaum Associates. [JSA]

Pennington, N., and Kosslyn, S. M. (in preparation) Measuring the acuity of the mind's eye. [SMK]

Pinker, S. (1979) The representation of three-dimensional space in mental images. Ph. D. thesis, Harvard University. [SMK]

(in press) Mental imagery and the third dimension. *Journal of Experimental Psychology: General.* [SMK]

Pinker, S., and Finke, R. A. (in press) Emergent two-dimensional patterns in images rotated in depth. *Journal of Experimental Psychology: Human Perception and Performance.* [SMK]

Pinker, S., and Kosslyn, S. M. (1978) The representation and manipulation of three-dimensional space in mental images. *Journal of Mental Imagery* 2:69–84. [SMK]

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

Pribram, K. H. (1969), The neurophysiology of remembering. *Scientific American* 220:73–86. [DLW]

Putnam, H. (1973) Reductionism and the nature of psychology. *Cognition* 2:131–46. [SMK, WES]

Pylyshyn, Z. W. (1973) What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychological Bulletin* 80:1–24. [SMK, BS, PWS]

(1978) Foundations of cognitive science. Presentation to the New Harvard Center for Cognitive Studies, Cambridge, Mass. [SMK]

(1979) The rotation of mental images: a test of the "holistic analogue" hypothesis. *Memory and Cognition* 7:19-28. [JMK, SMK, BS]

(1980a) Computation and cognition: issues in the foundations of cognitive science. *The Behavioral and Brain Sciences,* 3 (1). [ZP]

(1980b) The imagery debate: analogue media or tacit knowledge? (submitted for publication). [ZP]

(in press) Validating computational models: a critique of Anderson's indeterminacy of representation claim. *Psychological Review.* [JMK, SMK]

Reed, S. K. (1974) Structural descriptions and the limitations of visual images. *Memory and Cognition* 2:329–36. [GH]

Richman, C. L.; Mitchell, D. B.; and Reznick, J. S. (1979) Mental travel: some reservations. *Journal of Experimental Psychology: Human Perception and Performance* 5:13–18. [SMK, CLR, PWS]

Rock, I. (1973) *Orientation and form.* New York: Academic Press. [GH]

Roe, A. (1951), A study of imagery in research scientists *Journal of Personality* 19:459–70. [AR]

Rosenthal, R., and Rosnow, R. L., eds. (1969) *Artifact in behavioral research.* New York: Academic Press. [SMK]

Ryle, G. (1949) *The concept of mind.* London: Hutchinson & Co. [SMK]

Sarbin, T. (1972) Imagining as muted role-taking: a historical-linguistic analysis. In *The function and nature of imagery,* ed. P. W. Sheehan. New York: Academic Press. [AR]

Schank, R. C., and Abelson, R. P. (1977) *Scripts, plans, goals and understanding: an inquiry into human knowledge structures.* Hillsdale, N.J.: Erlbaum Associates. [RPA, DLW]

Segal, S. J., and Fusella, V. (1970) Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology* 83: 458–64. [SMK]

Sekuler, R., and Nash, D (1972) Speed of size scaling in human vision. *Psychonomic Science* 27:93-94. [SMK]

Seuss, B. (1957) *The Cat in the Hat Comes Back.* New York: Random House. [SMK]

Sheehan, P. W. (1967) A shortened form of Betts' Questionnaire upon mental imagery. *Journal of Clinical Psychology* 23:286–89. [AR]

(1978) Mental imagery. In *Psychology survery, No. 1* ed. B. M. Foss, London: Allen & Unwin. pp. 58–70. [PWS]

Shepard, R. N. (1978) The mental image. *American Psychologist* 33:123–37. [SMK, AR]

Shepard, R. N., and Feng, C. (1972) A chronometric study of mental paperfolding. *Cognitive Psychology* 3:228–43. [SMK]

Shepard, R. N., and Judd, S. A. (1976) Perceptual illusion of rotation of three-dimensional objects. *Science* 191:952–54. [SMK]

Shepard, R. N., and Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science* 171:701–03. [GH, SMK, TPM, BS, DLW]

Shepard, R. N. and Podgorny, P. (1978) Cognitive processes that resemble perceptual processes. In *Handbook of learning and cognitive processes,* vol. 5 ed. W. K. Estes. Hillsdale, N.J.: Erlbaum, Associates. [SMK]

Shulman, G. L.; Remington, R. W.; and McLean, J. P. (1979) Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance* 5:522–526. [SMK]

Schwartz, S. P. (1979) Studies of mental image rotation: implications of a computer simulation model of visual imagery. Ph.D. thesis, Johns Hopkins University. [SMK]

Simon, H. A. (1972) What is visual imagery? An information processing interpretation. In *Cognition in learning and memory,* ed. L.W. Gregg. New York: John Wiley. [SMK]

Simon, H. A. (1969) The architecture of complexity. In *The Sciences of the Artificial,* ed. H. A. Simon. Cambridge, Mass.: M.I.T. Press.

Singer, G., and Sheehan, P. W. (1965) The effect of demand characteristics on the figural after-effect with real and imaged inducing figures. *American Journal of Psychology* 78:96–101. [PWS]

Singer, W. (1979) Temporal aspects of subcortical contrast processing. *Neurosciences Research Program Bulletin* 15:358–69. [BB]

Smith, E. E. and Nielson, G. D. (1970) Representation and retrieval processes in short-term memory: recognition and recall of faces. *Journal of Experimental Psychology* 85:397–405. [SMK]

Smith, E. E.; Shoben, E. J.; and Rips, L. J. (1974) Structure and process in semantic memory: a feature model for semantic decisions. *Psychological Review* 81:214–41. [SMK]

Spinelli, D. N., and Pribram, K. H. (1967) Changes in visual recovery functions and unit activity produced by frontal and temporal cortex stimulation. *Electroencephalography and Clinical Neurophysiology* 22:143–49. [BB]

Spinelli, D. N.; Pribram, K. H.; and Weingarten, M. (1965) Centrifugal optic nerve responses evoked by auditory and somatic stimulation. *Experimental Neurology* 12:303–19. [BB]

Spinelli, D. N., and Weingarten, M. (1966) Afferent and efferent activity in single units of the cat's optic nerve. *Experimental Neurology* 13:347–61. [BB]

Spoehr, K. T., and Williams, B. E. (1978) Retrieving distance and location information from mental maps. Paper presented at the 19th annual meeting of the Psychonomic Society, San Antonio. November [SMK]

Sussman, G. J. (1973) The FINDSPACE problem (memo 286) Cambridge, Mass.: MIT A.I. Lab. [DLW]

Treyens, J. C. and Brewer, W. F. (1978) The effects of expected probability and expected saliency on memory for objects in a room. Paper presented at Annual Meeting of the MPA, Chicago. [DLW]

Waltz, D. L. (1979) Relating images, concepts, and words. Proceedings of the NSF Workshop on the Representation of 3-D Objects, University of Pennsylvania, 1-29. [DLW]

Waltz, D. L., and Boggess, L. C. (1979) Visual analog representations for natural language understanding, to appear in Proceedings of the Sixth International Joint Conference on Artificial Intelligence, Tokyo. [DLW]

Weber, R. J., and Harnish, R. (1974) Visual imagery for words: the Hebb test. *Journal of Experimental Psychology* 102:409–14. [SMK]

Weber, R. J.; Kelley, J.; and Little, S. (1972) Is visual imagery sequencing under verbal control? *Journal of Experimental Psychology* 96:354–62. [SMK]

Willis, J. (1621) *The art of memory as it dependeth upon places and ideas.* London. Facsimile published by Da Capo Press, New York, 1973. [IMLH]

Wilton, R. N. (1978) Explaining imaginal inference by operations in a propositional format. *Perception* 7:563–74. [CLR]

Yates, F. A. (1966) *The art of memory.* London; Routledge and Kegan Paul. [IMLH]