Psychological Review
1981, Vol. 88, No. 1, 46–66

0033-295X/81/8801-0046$00.75

The Medium and the Message in Mental Imagery: A Theory

Stephen Michael Kosslyn
Harvard University

A computational theory of imagery is described in this article. This theory posits that visual mental images are transitory data structures that occur in an analogue spatial medium. These "surface" representations are generated from more abstract "deep" representations in long-term memory and, once formed, can be operated upon in various ways. The theory is described in terms of detailed claims about the mental structures and processes invoked during imagery. In addition, the philosophical and empirical roots of the present theory are briefly reviewed. Further, arguments and data that have been offered against the theory are critically examined, and none are found damaging. An alternative account of the data that purportedly support the theory is also examined and found deficient in several respects. Finally, the current status of the "analogue–propositional" debate is reviewed, and it is concluded that there has been genuine progress in our understanding of the issues during the past decade.

In the *Theaetetus* Plato likened memory representations to impressions on a wax tablet, perhaps thereby becoming the first theorist to distinguish between representations (the different possible impressions) and the medium in which they occur (the wax tablet). The distinction between a representation and a medium has proven important in the study of visual mental imagery. Although no serious researcher today maintains that images are actual pictures in the head, some still find it reasonable to posit quasi-pictorial representations that are supported by a medium that mimics a coordinate space. On this view, images are not languagelike "symbolic" representations but bear a nonarbitrary correspondence to the thing being represented. Partly because of the primitive origins of this idea, many people seem wary of it. But the idea that images are a special kind of representation that depicts information and occurs in a spatial medium is not patently ridiculous, and in fact can be developed in a very coherent way that violates neither philosophical nor empirical considerations. In this article I will sketch out one way this is being done and will show that none of the criticisms of this approach to understanding mental imagery, recent (especially see Pylyshyn, 1979a, 1980, 1981) or traditional (see Kosslyn & Pomerantz, 1977), are penetrating or incisive. Further, I will also show that none of the data that purportedly speak against this approach are in fact damaging. Finally, I will consider the relative merits of two kinds of accounts of imagery data: those based on processing of depictive images and those based on appeal to the influence of demand characteristics, task demands, and use of tacit knowledge.

I wish to thank Ned Block, Sharon Fliegel, Alexander George, Reid Hastie, Edward E. Smith, George E. Smith, and Lucia Vaina for insightful comments and illuminating criticisms of an earlier draft of this article. Work reported herein was supported by National Science Foundation Grant BNS 79-12418.

Requests for reprints should be sent to Stephen M. Kosslyn, Department of Psychology and Social Relations, Harvard University, 1236 William James Hall, 33 Kirkland Street, Cambridge, Massachusetts 02138.

Background Assumptions

Before one begins to theorize one should have a reasonably clear conception of both the domain of the theory and the form the theory should take. In addition, I have found it useful and important to distinguish between the theory proper and two sorts of models, specific and general.

The Domain of the Theory

The goal of this article is to describe a theory of how information is represented in,

and accessed from, visual mental images. For example, when asked to count the number of windows in their living room, most people report mentally picturing the walls, scanning over them, and "looking" for windows. The present theory is intended to provide accounts of this "mental picturing" process, of "looking" at images, and of transforming images in various ways. In addition, the theory also specifies when images will be used spontaneously in the retrieval of information from memory (as in the foregoing example).

The Form of a Cognitive Theory of Imagery

A cognitive account of imagery is a theory about the *functional capacities* of the brain—the things it can do—that are invoked during imagery. There are numerous ways to describe the range and kinds of functional capacities involved in any given domain of processing, but most theorists have found it useful to describe these capacities in terms of structures and processes. Let us distinguish between two kinds of structures, *data structures* and *media*, and two general kinds of processes, *comparisons* and *transformations*.

Data structures. Data structures are the information-bearing representations in any processing system. They can be specified by reference to three properties, their *format*, *content*, and *organization*. The format is determined by (a) the nature of the "marks" used in the representation (such as ink, magnetic fluxes, or sound waves) and (b) the way these marks are interpreted (the mark *A* could be taken as a token of a letter of the alphabet or a picture of a particular pattern). The format specifies whether a representation is composed of primitive elements and relations and, if so, specifies their nature. The content is the information stored in a given data structure. Any given content can be represented using any number of formats. For example, the information in the previous sentence could be stored on a magnetic tape, on a page, as a series of dots and dashes etched on metal, and so on. The organization is the way elementary representations can be combined. The format of a representation constrains the possible organizations but does not determine them. For example, propositional representations can be ordered into various kinds of lists and networks.

Media. A medium does not carry information in its own right. Rather, a medium is a structure that supports particular kinds of data structures. This page, a TV screen, and even the air are media—supporting ink, glowing phosphor, and sound patterns, respectively. Media can be specified by reference to their *formatting* and *accessibility*. The formatting places restrictions on what sorts of data structures can be supported by a medium. A short-term store, for example, might have five "slots" that take "verbal chunks"—but not visual images or abstract propositions. The accessibility characteristics dictate how processes can access data structures within a medium. The slots of a short-term store, for example, might be accessible only in a given sequence.

Note that all of the properties of the media and the data structures are by necessity defined in the context of a particular processing system. Even though structures have an independent existence, and their nature imposes constraints on the kinds of processes that can be used (see Hayes-Roth, 1979; Keenan & Moore, 1979; Pylyshyn, 1979c), structures attain their functional properties only vis-à-vis the operation of particular processes. For example, if items on a list can be retrieved only in one order on one day and another order on the next day, the functional order of the list has changed—even though the data structure has not.

Comparison processes. These procedures compare two data structures or parts thereof and return a match/mismatch decision or a measure of the degree of similarity (defined over a specific metric) between the representations.

Transformation processes. There are two very general classes of transformation processes, *alterations* and *productions*. Alteration transformations operate to alter a given data structure by changing its contents (e.g., by adding or deleting an item on a list) or reorganizing it (e.g., by reordering items on a list). Production transformations, in contrast, leave the initial data structure in-

tact but use it as an impetus either to replace or to supplement it with a new data structure. This new data structure may differ from the initial one in its format (as when a pattern is described), in content (as when an initial image is replaced by one with more details), and/or in organization (as when a list is replaced by a new one with the same items but in a different order). It is difficult for me to conceive of how an alteration transformation can itself change the format of a data structure, and this may prove to be a critical distinction between the two classes (see chapter 5 of Kosslyn, 1980).

The reader should note that the actual expression of the theory may not preserve the individual functional capacities as distinct terms. 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. I make no commitment as to the form of such an ultimate abstract expression but only claim that it will express lawful relations among the kinds of cognitive entities described above. The job for empirical research programs at this time, as I see it, is to isolate and develop the clearest possible characterization of the individual functional capacities and their interrelations.

Specific Models and General Models

One way to begin to formulate a cognitive theory is to develop a model of the presumed functional capacities. Models differ from theories in at least two ways. First, models have three sorts of components: those that are theory-relevant, those that are not theory-relevant (e.g., in the case of a computer model, those aspects that are a consequence of the particular hardware being used), and those that are theory-neutral (that cannot be assigned to either of the other two classes with certainty; see Hesse, 1963; Kosslyn, Pinker, Smith, and Shwartz, 1979). Second, models contain an element of "as if" that is not present in a theory proper. That is, a model is assumed to be under a description, or under a certain interpretation, that leads one to draw points of similarity between it and the modeled domain. A theory proper is unambiguous and not in need of such interpretation.

It is useful to distinguish two basic kinds of models, *specific* and *general*. Specific models are designed to account for performance in a particular task (see Clark & Chase, 1972; Sternberg, 1969, for examples), whereas general ones embody the entire set of principles (assumptions about functional capacities and their interrelations) that should account for performance in all the tasks in a given domain (Anderson & Bower, 1973, and Newell & Simon, 1972, were developing general models; in physics, Bohr's atom was a general model). One problem with attempting to develop isolated specific models for particular tasks 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 specific 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 is forced to be consistent across tasks. If the rules of combination are specified precisely enough (and these are implicit in the input and output specifications of each process; the output from one usually will serve as the input to another), then a given particular input configuration will evoke only one sequence of operations, providing a specific model of how a particular task is accomplished. Thus, I assume that although a given task logically could be accomplished in more than one way and in fact may be done differently on different occasions (such as when one is tired versus 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.

The general model of image representation and processing we have developed takes the form of a computer simulation (see Kosslyn, 1980; Kosslyn & Shwartz, 1977, 1978). Each process is represented as a subroutine or set of subroutines, and each structure has been implemented as well (as described below). There are numerous virtues in building a general model of the sort we have been developing, but two stand out: First, if one

tries to motivate the decisions necessary to model an entire domain in a consistent, precise way, one will be inspired to collect new and interesting data to select among plausible alternative ways of building the model. Second, the model allows one to formulate precise accounts for performance in numerous specific tasks, and these accounts are self-consistent. Thus, one can test the theory as a whole by generating predictions about what should happen in particular tasks, as we in fact have done (see Kosslyn, 1980). On our view, then, the main reason to formulate a model for a particular task is to test the underlying assumptions of the theory that dictated that specific model. It is simply too easy to explain performance in any given task in isolation for this exercise to be very useful (as should be evident later in this article); it is only when one is trying to provide accounts for all the tasks in a domain that one seems to learn very much.

Overview of the Theory

The following review outlines the most central claims of the Kosslyn and Shwartz (1977, 1978; Kosslyn, 1980) theory of mental image representation and processing, which will be described in terms of the kinds of structures and processes discussed in the previous section. Properties of the general model embodying the theory will also be described occasionally to clarify the theoretical claims; unless otherwise noted, only the theory-relevant properties of our simulation will be mentioned (see Kosslyn, 1980, for a more detailed treatment). Although a few of the criticisms of this type of theory—and our particular theory per se—will be addressed in this section when relevant, a more thorough defense will be deferred until after the theory as a whole is sketched out.

Structures

On our view, images have two major components. The "surface representation" is a quasi-pictorial representation that occurs in a spatial medium; this representation depicts an object or scene and underlies the experience of imagery. The "deep representation" is the information in long-term memory that is used to generate a surface representation.

The Surface Representation

The properties of the surface image are in part a consequence of the properties of the medium in which it occurs, which we call the *visual buffer*. The visual buffer is implemented as an array in our computer simulation (the *surface matrix*); a surface image is represented by a configuration of points in this array that depicts an object or objects.[1]

1. The Medium

Formatting. (a) The visual buffer functions as if it were a coordinate space. This "space" is not an actual physical one but is rather a functional space defined by the way processes access the structure. The functional relations of the loci in the visual buffer need not be determined by actual physical relations any more than the functional relations of cells in an array in a computer need be determined by the physical relations among the parts of core memory. That is, the processes that operate on an image access the medium in such a way that local regions are separated from each other by different numbers of locations (i.e., differences in the number of just-noticeable differences in position in the coordinate space). We posit that the organization of the visual buffer is innately determined and fixed. Information is represented by selectively activating local regions of the space. (b) The visual buffer has a limited extent and specific shape, as measured empirically (see Finke & Kosslyn, 1980; Kosslyn, 1978), and hence can support only representations depicting a limited visual arc. This makes sense if this medium is also used in perceptual

[1] The fact that the array used to simulate the visual buffer is square is a good example of a property of the model that was not intended to be theory relevant. We never have claimed that the visual buffer is a strict Cartesian space with anisotropic properties resulting from a rigid organization into rows and columns. It is an empirical question whether this incidental feature of the model should be taken seriously or not; it is not a priori obvious, at least to me, that the spatial medium must be isotropic.

processing; if so, then it presumably only needed to evolve to represent input from the limited arc subtended by the eyes.

Accessibility. (a) The visual buffer has a grain, resulting in a limited resolution. Thus, portions of subjectively smaller images (i.e., those which seem to subtend a smaller angle) are more difficult to classify because subtle variations in contour are obscured (see Kosslyn, 1975, 1976a; an initial attempt at measuring the resolution of the medium is described in Pennington & Kosslyn, Note 1). (b) The resolution is highest at the center of the visual buffer and decreases toward the periphery (see Finke & Kosslyn, 1980; Kosslyn, 1978). Importantly, although grain is not homogeneous throughout the medium, at any given location it is presumed to be fixed. (c) Representations within the visual buffer are transient and begin to decay as soon as they are activated. This property results in the medium's having a capacity defined by the speed with which parts can be generated and the speed with which they fade; if too many parts are imaged, the ones activated initially will no longer be available by the time the later ones have been imaged. This property was posited in order to explain my finding (Kosslyn, 1975) that images of objects in complex scenes were more degraded than images of objects in simple contexts.

2. *The Data Structure*

Format. The surface image *depicts* an object or scene. The primary characteristic of representations in this format is that every portion of the representation must correspond to a portion of the object such that the relative interportion distances on the object are preserved by the distances among the corresponding portions of the representation (cf. Shepard, 1975). Three implications of this characterization are that (a) size, shape, orientation, and location information are not independent in this format—in order to depict one, values on the other dimensions must be specified; (b) any part of a depictive representation is a representation of a part of the represented object; and (c) the symbols used in a depiction (such as points in an array) cannot be arbitrarily assigned their roles in the representation (i.e., a given point must represent a given portion of the object or scene once the mapping function from image to object is established; on our view this function is innately determined and fixed by the human visual system). None of these properties are shared by discursive propositional (or "symbolic") representations (see chapter 3 of Kosslyn, 1980, for a detailed development of these points).

Thus, surface images consist of regions of activation in the visual buffer that correspond to regions of depicted objects, with distances among the regions on an object (as seen from a particular point of view) being preserved by distances among the regions used to represent it in the medium. Importantly, *distance* in the medium can be defined without reference to actual physical distance but merely in terms of the number of locations intervening between any two locations.

It is important to note that when terms such as *distance* and *orientation* are used to refer to surface images, they are being used in a technical way, referring to functional relations among regions in the visual buffer. Increased distance, for example, will be represented by increased numbers of locations in the visual buffer. Thus, although there is no physical distance or orientation in a depictive representation in the visual buffer, the corresponding states can sensibly be interpreted by using these terms. Contrary to what Pylyshyn (1981) asserts, we have not committed an erroneous "slip of scope" by talking about properties of the image rather than properties of the imaged object. In perception one can talk about properties of the "optical array" (such as those noticed by painters who use perspective) and of the objects themselves; whereas the size of an object does not change with distance, its "size" in the optical array (angle subtended) does. Similarly, in imagery we can speak of properties of the image itself by reference to the position, location, area occluded, and so on in the visual buffer (the analogue spatial medium). In this case it makes no difference what the image is an image of; the "subjective size" is independent of the actual size of the object. And in point of fact, a number

of processes—such as scanning (see Kosslyn, Ball, & Reiser, 1978)—depend on the subjective size, not on the actual size of objects.

Content. Images depict appearances of objects seen from a particular point of view (and hence are *viewer-centered*, to use the term of Marr and Nishihara, 1978). Images may represent the actual objects depicted, or images of objects may be used to represent other information (as occurs, for example, if one represents the relative intelligences of three people by imaging a line with three dots on it, a dot for each person; see Huttenlocher, 1968). Note that the content is determined not just by the image itself but also by how the interpretive processes "read" the image. We are specifying the way a system of representations and processes operates in which the properties of the components are to some degree mutually interdependent.

Organization. Individual images may be organized into a single composite representing a detailed rendition of a single object or a scene. Because parts of images are theorized to be generated sequentially, and parts begin to fade as soon as they are imaged, different parts of the image will be at different levels of activation. Level of activation ("fade phase") will dictate an organization of the surface image because points at the same level will be grouped together according to the Gestalt Law of Common Fate.

The Underlying Deep Representations

Our findings suggest that there are two types of representations in long-term memory that can be used to generate images, which we call *literal* and *propositional*. Literal information consists of encodings of how something looked, not what it looked like; an image can be generated merely by activating an underlying literal encoding. Propositional information describes an object, scene, or aspect thereof and can be used to juxtapose depictive representations in different spatial relations in the visual buffer.

The Literal Encodings

1. The medium

Formatting. The long-term memory medium does not function as if it were a coordinate space. Rather, it stores information in nonspatial units analogous to the files stored on a computer. In our computer simulation model, files store lists of coordinates specifying where points should be placed in the surface matrix to depict the represented object or objects (but we do not theorize that images are sets of dots or that underlying literal encodings are sets of coordinate pairs; these implementation details are not meant to be theory-relevant). The units are identified by name.

Accessibility. The units are accessed by name; the extent of the represented object along a given dimension can be computed without first generating a surface image; and the representation can be sampled a portion at a time. (These last two properties were posited in order to explain how parts of objects can be imaged at the appropriate size on a foundation part; for details see chapters 5 and 9, Kosslyn, 1980).

2. The data structure

Format. We have not as yet made any strong claims about the precise format of the underlying literal encodings.

Content. The underlying literal encodings have the same content as the surface images they can produce.

Organization. Every object is represented by a "skeletal encoding," which represents the global shape or central part. In addition to the skeletal encoding, objects may be represented by additional encodings of local regions or parts. (See chapters 4, 6, and 7 of Kosslyn, 1980, for data that bear on these claims.) Multiple encodings are linked by propositional relations that specify where a part belongs relative to another part or the skeleton. (This property seemed necessary to explain the flexibility with which images may be reorganized and combined in accordance with a new description; see chapters 4 and 6 of Kosslyn, 1980.)

The Propositional Encodings

1. The medium

Formatting. The medium is structured to contain ordered lists of propositions, and the lists are named.

Accessibility. Lists are accessed by name, and are searched serially, starting from the "top." (This assumption allowed us to explain the effects of association strength on property-verification times and led to some interesting predictions about image generation; see chapter 7, Kosslyn, 1980.)

2. *The data structure*

Format. The entries in these lists are in a propositional format. Propositions are abstract languagelike discursive representations, corresponding roughly to simple active declarative statements. Kosslyn (1980) presents a more detailed and formal treatment of propositional representation, but this general characterization is sufficient for present purposes.

Content. Lists contain information about (a) parts an object has (included in order to explain how detailed images can be generated and in order to model question-answering processes); (b) the location of a part on an object (necessary in order to integrate separate encodings into a single image); (c) the size category of a part or object (necessary in order to adjust the size scale so that a part or object will be optimally resolved); (d) an abstract description of a part or object's appearance (required for the interpretive processes to identify the pattern of points depicting a part or object); (e) the name of the object's superordinate category (included for inference procedures used during question answering); and (f) the name of literal encodings of the appearance of the object (necessary in order to integrate multiple encodings into a single image).

Organization. Pointers in lists indicate which other list or lists to look up in sequence, resulting in lists' being organized hierarchically or in any graph structure.

Processes

The imagery theory at present provides accounts for four classes of imagery processing: those involved in image generation, inspection, or transformation, and those that determine when imagery will be used spontaneously when people retrieve information from long-term memory. We have also begun to extend the theory to answer questions about how images are encoded as mnemonic devices and the role of imagery in reasoning. Space limitations preclude a detailed description of the processing components we posit or of how they are invoked when one is performing a specific task (see Kosslyn, 1980). In brief, the major processing components are as follows.

Image Generation

Image generation occurs when a surface image (which is quasi-pictorial) is formed in the visual buffer on the basis of information stored in long-term memory. Image generation is accomplished by four processing components, which we call PICTURE, FIND, PUT, and IMAGE. The PICTURE process converts information encoded in an underlying literal encoding into a surface image (in the model, it prints points in the cells of the surface matrix specified by the coordinate pairs stored in the underlying literal file). The PICTURE process can map the underlying representation into the visual buffer at different sizes and locations, depending on the values of the size and location parameters given it. (This property was motivated by our finding that people can voluntarily form images of objects at different sizes and locations; see chapter 4 of Kosslyn, 1980.) The FIND process looks up a description of an object or part and searches the visual buffer for a spatial configuration that depicts that object or part. This process is used when multiple literal encodings are amalgamated to form a single image in the visual buffer; in this case the FIND process locates the "foundation part" where a new part should be added to previously imaged material. The PUT process performs a variety of functions necessary to image a part at the correct location on an image, including looking up the location relation in the list of propositions associated with an object and adjusting the size of the to-be-imaged part. The IMAGE process coordinates the other processing components and, in so doing, determines whether an image will be detailed (i.e., include parts stored in separate literal encodings) or undetailed (i.e., be constructed solely on the basis of the skeletal encoding).

The IMAGE process is invoked by a command to form an image of a given object, either detailed or not, at some specified size and location (or at a default size and location). All of the processes used in image generation are production transformations except FIND, which is a comparison process.

Image Inspection

Image inspection occurs when one is asked a question such as, "Which is higher off the ground, the tip of a racing horse's tail or its rear knees?" and one "looks at" an image of the horse with the "mind's eye." The process of "looking" is explained by reference to a number of distinct processing components, notably LOOKFOR (a production transformation). The LOOKFOR process retrieves the description of a sought part or object, looks up its size, employs the RESOLUTION process (a production transformation) to determine if the image is at the correct scale, adjusts the scale if need be by invoking the ZOOM or PAN process (alteration transformations), scans to the correct location, if necessary, by using the SCAN process (also an alteration transformation), and then employs FIND to search for the sought part. If the sought part is not found, the PUT process is used to elaborate the image further by generating images of parts that belong in the relevant region, and then FIND is used to inspect the image again. Note that because the FIND process is used in both image inspection and image generation, we should discover effects of the ease of executing this process in both kinds of tasks. And sure enough, less discriminable parts are not only more difficult to "see" during image inspection but are more difficult to locate as foundation parts (i.e., places where additional parts will be placed) during image construction (see chapters 4 and 6 in Kosslyn, 1980).

Image Transformation

According to our theory there are two classes of transformations and two modes of performing these transformations.

Classes of transformations: Field general and region bounded. Field-general (FG) transformations alter the entire contents of the medium, of the visual buffer, without respect to what is actually represented. Region-bounded (RB) transformations first delineate a region in the visual buffer and then operate only within the confines of that region. Virtually every FG transform has an RB analogue. For example, "zooming in" is FG but "growth" is RB. According to our theory, the number of objects manipulated should affect processing time only for RB transformations, because each object is manipulated separately here (but not in the FG case). Pinker and Kosslyn (1978) present some data that support this distinction (although this was not realized at the time the experiment was conducted; see Kosslyn et al., 1979).

Modes of transformation: Shifts and blinks. The bulk of the data on image transformations suggest that images are transformed incrementally, passing through intermediate points along a trajectory as the orientation, size, or location is altered (see chapter 8, Kosslyn, 1980). This property is a hallmark of *shift* transformations, which operate by translating the locations of individual portions of the data structure to new locations in the visual buffer (and it remains an empirical question how to define *portion*). Because the system is inherently noisy (as are all physical systems), if portions are moved too far, they become too scrambled to be realigned by "cleanup routines." Thus, the limits of the cleanup routines force the processor to translate points in a series of relatively small increments. This results in greater transformations requiring more operations and hence more time. Note that this account hinges on the amount of distortion increasing as points are translated greater "distances"; if this were not true then portions could be "moved" the entire "distance" in one increment. The idea of increases in distortion with larger step sizes makes sense if the transformation process makes use of an analogue adder and multiplier in which the bigger the value, the larger the range of error. Further, because scanning is treated as another form of image transformation (in which a SCAN process shifts the data structure through the visual buffer, so that different parts fall in the center of the medium and hence are most sharply in focus),

the same principles apply to it as to other forms of image transformation (such as rotation). The ZOOM and PAN processes dilate and contract the image, respectively, and the ROTATE process alters the orientation of the image. All "sizes" and "orientations" are, of course, defined relative to the visual buffer. SCAN, ZOOM, and PAN are field general and ROTATE is region bounded.

We did not initially plan on positing a second mode of image transformation. However, we were stuck with the possibility of *blink* transformations, given our prior claims that surface images can be generated (from the underlying deep representations) at optional sizes and locations and that images fade over time. Given these assumptions, we were forced to assume that people can transform images by letting an initial image fade and then generating another image of the object at a new size or location (and hence the contents or organization of an image can be changed via a production transformation as well as via an alteration transformation). In this case the transformation is discontinuous; images do not pass through intermediate states of transformation. The reason shift transformations are the default, we claim, is that they generally are less effortful (i.e., fewer and less complex operations are required to manipulate an existing image than to generate a new one from the underlying representations). But in the case of shift transformations, effort increases with the extent of the transformation (because more iterations are required)—but not so with blink transformations. Thus, there will come a point when it is "cheaper" to abandon the initial image and generate a new one. We have in fact collected data supporting this claim (see chapter 8, Kosslyn, 1980).

But how do people know which mode of transformation will be more efficient before using one? In scanning, for example, it is possible that people can decide which transformation to use on the basis of an initial estimate of the distance to be scanned, which can be computed in any number of ways, for example by using the underlying propositional representations of location (which we needed to posit in order to explain how parts can be placed at the correct location on an image during the generation process). In addition to explaining data, the distinction between a shift and a blink transformation has led us to make a number of predictions, some of which are not intuitively obvious (see Kosslyn, 1980).

Spontaneous Use of Imagery in Fact Retrieval

Imagery is likely to be used in fact retrieval if the fact is about a visible property of an object a person has seen and it has not been considered frequently in the past. According to our theory, image encodings are accessed in parallel with propositional ones. Thus, the more overlearned the propositional information is (and hence the higher the entry on an object's propositional list, according to our theory), the more likely it is that a propositional encoding will be looked up or deduced before image processing is complete (i.e., before an image can be generated and inspected). Thus, imagery will often be used in retrieving information about objects learned "incidentally," as occurs when one retrieves from memory the number of windows in a room or considers the shape of a dog's ears. In addition, images—by the very nature of the format—make explicit information about relative shapes, relative positions of objects or parts thereof, and the appearances of objects and parts as seen from a particular point of view. Thus, when these sorts of information are required in order to make a judgment, imagery will often be used. However, because most objects are categorized, at least roughly, along these kinds of dimensions in a presumably propositional format (which may be well learned and hence likely to be accessed prior to image processing), imagery is most likely to be used when relatively subtle comparisons along these dimensions are required (and the to-be-compared objects fall in the same propositionally encoded category, preventing such category information from being used to make a judgment). For example, imagery should be used if one is asked to decide which is larger, a hamster or a mouse (both presumably are categorized as "small"), but should not be used in deciding which is larger, an elephant or a rabbit (which presumably fall in different categories). These

principles are derived from results presented in Kosslyn, Murphy, Bemesderfer, and Feinstein (1977), Kosslyn and Jolicoeur (1980), and Kosslyn (1980, chapter 9).

Criticisms of the Theory

Pylyshyn (1979a, 1980, 1981) offers numerous specific criticisms of our theory. These criticisms are of two sorts, purported deficiencies of our theory itself and the relative virtues of an alternative theory.

Specific Criticisms of the Kosslyn and Shwartz Theory

The five most important criticisms of our theory are discussed below.

Ad Hoc Theories

Pylyshyn seems to subscribe to the view that if the properties of theoretical entities are not constrained a priori, they are equivalent to "free empirical parameters" that are merely stipulated ad hoc in order to explain data. The broad form of virtually any theory is guided by a wide range of considerations (e.g., shared mechanisms with perception, a clear mechanism for ontogenesis, and possible instantiation in the brain, in our case—see Kosslyn, 1980, chapters 5 and 10, and Kosslyn & Shwartz, 1978). But where are the specific claims of a theory supposed to come from, if not from an attempt to explain data in a perspicuous fashion? It is true that the data that provided the motivation to adopt a specific theoretical position are explained ad hoc by that part of the theory, but this does not mean that the theory is ad hoc: When new data are explained, and these new data are not similar to the old data that provided the initial motivation for formulating the theory, this explanation is clearly not ad hoc. Importantly, once we characterize a property we do not casually revise our theory simply to explain additional data.

Parameters and Free Parameters

Pylyshyn also decries the number of components specified by the theory, suggesting that we are in danger of having an unlimited number of free parameters. We must distinguish between number of parameters and number of free parameters. It simply is not true that we are in danger of having as many free parameters as we have components, as should be evident even in the brief description of the theory offered above. The parameters (properties of the theory) are not left vague enough to adjust whenever the whim comes over us. Properties of the medium, the data structures, and the processing components we posit place genuine constraints on the kinds of explanations we can formulate. For example, an important feature of the theory is our assumption that the medium does have a fixed grain (which we have begun to measure; see Pennington & Kosslyn, Note 1) that places an upper bound on the degree of resolution with which one can image a given object. That is, one can image an object at less than optimal resolution—just as one can draw, paint, or represent an object on photographic film with less than optimal resolution—but one cannot image at greater resolution than the medium will allow. The degree to which the theory is in fact well specified should be apparent in the efforts necessary to provide detailed accounts for some of the data reviewed in Kosslyn (1980).

Pylyshyn especially decries the fact that we theorize that the generation and transformation processes are flexible. Pylyshyn may be bothered by two things: missing details of the theory and optional procedures (e.g., an image can be generated with or without details). With regard to the first worry, indeed, we have not presented a complete theory. Many properties of the theory are currently open. The alternative would be to simply make up properties on the basis of our intuitions, which strikes us as overly optimistic at best and a meaningless exercise in fantasy at worst. With regard to the second complaint, we do in fact posit that some of the process components are flexible and alter their operation depending on the input. We see no reason why this need not be true—even though it makes things less convenient for the theorist.

The Source of Predictive Power

Pylyshyn has claimed that the predictive power of our theory is deceptive. One criti-

cism is that our imagery theory is really just a restatement of commonsense analogies to perception. First of all, it should be apparent even from the brief overview offered above that our theory is not based on a simple analogy to perception. Rather, we have tried to specify the nature of the mental structures and processes that are used in mental image representation and processing per se. Further, the predictions of the theory do not hinge on commonsense knowledge of perception or the world. Predictions from the field general/region bounded distinction, between transformations that operate over the visual buffer as a whole (regardless of its contents) and ones that operate only within a bounded region, are not intuitive. Nor is the claim that parts are often inserted into an image only at the time of inspection, and hence the association strength between an object and part should affect times in the same way in image inspection and image generation. (This prediction rests on claims about when and how ordered lists of propositions are searched; see chapters 5 and 7, Kosslyn, 1980.) Nor are the predictions about when imagery should be spontaneously used when people answer questions (see Kosslyn & Jolicoeur, 1980), and so on.

Pylyshyn has denied that our theory has genuine explanatory or predictive power on other grounds. He claims—and suggests that we are ready to admit—that much of our model's explanatory and predictive capacity derives from our original cathode ray tube metaphor of imagery (which merely likened images to spatial displays on a CRT screen that could be classified as depicting instances of a category and were generated from more abstract underlying representations). This simply is not true. The explanatory and predictive power of the theory and general model are drawn from the explicit claims we have made about the nature of internal structures and processes, which are not simple reworkings of the minimal assumptions that underlay the initial metaphor. Scanning is a case in point: In the original CRT metaphor (see Kosslyn, 1975), scanning was considered as shifting the point of focus across the image, whereas in the current theory, scanning is considered as a kind of image transformation in which the data structure is shifted across the visual buffer, with different portions falling in the most highly resolved, center region. The current theory allows us to explain why scanning times increase linearly at the same rate when one scans across the visible portions of the image as when one scans to a part that initially had "overflowed" the medium (see chapters 5 and 8 of Kosslyn, 1980). In addition, the theory predicts similarities between scanning and other image transformations (e.g., size scaling) that would never have been predicted on the basis of the CRT metaphor; for example, consider the properties of "blink scans," to be discussed shortly. The second half of Kosslyn's (1980) book presents numerous examples of explanations and predictions that clearly rest on more than the minimal assumptions underlying the original CRT metaphor.

Cognitive Penetration

Cognitive penetration occurs when one's knowledge, beliefs, goals, or other cognitive states alter performance in a "semantically systematic" way (see Pylyshyn, 1981), as occurs when one's knowledge of the laws of color mixing influences how colors appear in an image when they are mixed.[2] Cognitive penetration is important insofar as it demonstrates that properties of structures or processes are not fixed. We have claimed that the properties of the visual buffer are innately determined and that they should not be subject to cognitive penetration. As outlined below, one cannot take the existence of cognitive penetration in imagery tasks as evidence that this claim is incorrect.

Tasks and components. It has long been acknowledged in cognitive psychology that

[2] The actual technical definition of *cognitive penetration* is considerably more subtle than this one. The added subtlety is necessary for Pylyshyn to avoid classifying various noncognitive phenomena as cognitive by this criterion. For example, blood flow in the brain is affected by one's beliefs about stimuli one is viewing (because some stimuli can excite one, affecting heart rate and blood vessel dilation). But no one would want to classify blood flow as a cognitive process. It is not clear to me, however, that even Pylyshyn's more complex and subtle technical definition of cognitive penetration necessarily excludes cases like this.

performance of any given task involves numerous processing components. Not only are there distinct stages of processing, such as those used in encoding, in central processing of the encoded information, and in generation of responses, but each stage involves an interplay of structures and processes (see Anderson, 1978; Sternberg, 1969). Unlike the "additive factors" methodology developed by Sternberg (1969), which is designed to allow one to identify the effects of given variables with distinct processing stages, the cognitive penetrability criterion applies to tasks taken as a whole: Cognitive penetration demonstrates that the operation of at least one component of processing in a given task is systematically altered by the meaning of an input, but it does not serve to specify which component has been affected. If the effects of cognitive penetration are localized to processes that access a fixed analogue spatial medium, then the mere existence of cognitive penetration does not show that properties of this medium play no role in information processing—even if for some tasks one does not necessarily have to invoke these properties to provide one possible account for the data. Whether or not the properties of the medium constrain performance in any given task is an empirical question. At the present juncture it is important to note that the existence of nonanalogue components in a given set of processes in no way bears on the truth or falsity of the claim that one component is an analogue spatial medium which supports mental images that depict an object or scene.

The necessary and the typical. All parties agree that a cognitive theory ought to specify the constraints imposed on processing by a particular structure recruited during processing. This is to be distinguished from the characteristics of processing in a habitual or typical way. However, in order to know what behavioral consequences *must* be incumbent on processing a given structure, one needs to know more than just the properties of the structure itself: One also needs to know the properties of the process. It is important to realize that representations in an analogue medium need not be operated on solely by analogue processes. For example, we theorize that the visual buffer is accessed both by analogue processes, such as those involved in performing shift transformations, and by nonanalogue processes, such as those involved in classifying a spatial pattern as a depiction of a particular object or part. Depending on which sort of process is accessing the medium, a given property of the medium may or may not be evident during task performance. For example, the distance between two images in the most resolved region of the visual buffer should affect the time to move them together but should not affect how vividly they appear when inspected. Thus, it is critical that we realize that the only way to discover the properties of any structure is to observe how numerous different processes operate on it and at the same time observe how these processes operate on different structures, trying to abstract out which characteristics of the behavior are a consequence of properties of the structure and which are a consequence of properties of the processes.

The necessity of prior knowledge. It seems clear that at least some imagery phenomena, such as imaging what color results when one mixes red and green, require a prior knowledge of the principles involved. However, Pylyshyn wants to maintain that most imagery phenomena depend on one's prior knowledge and offers an argument with the following structure: First, he presents two lists of imagery phenomena, ones that do not seem to exhibit the effects of cognitive penetration (i.e., the effects of prior knowledge altering the image to achieve the correct result) and ones that do. Next, he claims that there is no principled way of distinguishing between the two sets of phenomena. Therefore, we are to conclude that it is more plausible to assume that all are a consequence of nonanalogue processes. Consider the following two responses to this argument:

First, the lack of a ready principle does not mean there will not be one eventually. For example, most of the cases Pylyshyn cites that do not seem to require the influence of prior knowledge seem to depend on purely spatial properties of arrays containing relatively few units, whereas the others do not. The examples that cannot be characterized this way, namely, dropping an object

and watching it fall and throwing a ball and watching it bounce, do not seem autonomous in the same way as do the other examples (e.g., a dot imaged in a square remains in the square while its sides are expanded). And in fact these two examples seem misclassified to me, both seeming to depend on one's knowledge of physical law as opposed to geometrical properties of the image representation per se.

Second, the plausibility argument can cut either way: At least some of the examples Pylyshyn cites do not seem to depend on prior knowledge (e.g., the dot in the expanding square) and are plausibly interpreted as revealing properties of an analogue spatial medium. Given that even one phenomenon leads one to posit an analogue spatial medium, it then makes sense to make use of the putative properties of this medium in providing straightforward accounts for other data. Whether or not these accounts are correct is, of course, an empirical question.

Cognitive penetration and mental rotation. Pylyshyn has offered the results of two experiments he reported earlier (Pylyshyn, 1979b) as evidence that the analogue account of mental rotation is wrong. We must be careful here because Pylyshyn is using the term *analogue* to include the entire operation: Encoding, rotation, and comparison processes are all considered part of a single analogue process. I am not certain who in fact subscribes to this position, but even if one were to maintain such a view, Pylyshyn's results are not a cause for concern. In these experiments subjects saw two stimuli simultaneously, one a geometric figure (containing internal lines dividing it into subpatterns) and the other possibly a part of the figure. The part was presented at different angular disparities from the orientation of the figure. The subject was asked to decide whether the part was a component of the figure, rotating the figure into congruence with the part if necessary. As is usual with this sort of experiment, decision times increased with the amount of rotation required to compare the stimuli. The main result of interest here, however, is the rate of increase in time with the amount of rotation: The amount of time required with greater amounts of rotation increased more sharply when the part was not a "good" (in the Gestalt sense) subpattern. That is, "bad" parts were apparently rotated at a slower rate. Pylyshyn argues that because subjects' knowledge of part goodness affected rotation, rotation is not an analogue process. Given that the goodness of a part was apparent only after it was located in a figure, it is not clear how any model could account for such clairvoyant behavior—which leads me to suspect that subjects did not simply rotate figures (at different rates) until part of the figure matched or did not match the probe part.

In fact, Pylyshyn's (1979b) results may not have anything to do with "goodness" per se: Because Pylyshyn used different patterns for his "good" and "bad" parts (instead of constructing sets of figures such that the same pattern could serve in both conditions), we cannot know whether this result is due to peculiarities of the individual patterns per se or to the relationship between the pattern and the figure in which it was embedded. But in any case, Pylyshyn's results may simply reflect task-specific strategies that have nothing to do with mental rotation as it occurs when stimuli are not presented simultaneously (as in the experiments reported by Cooper & Shepard, 1973) and hence are not available for successive visual comparisons. For example, they may reflect, as Pylyshyn himself notes, a "piecemeal rotate and compare" (p. 27) process. Perhaps the subjects did not encode the entire figure into a mental image but encoded only parts that they hoped would help in performing the task. If they guessed wrong, they fixated again on the figure and re-parsed it, encoding different parts into the image. In this case, when the test part corresponded to a "bad" part of the figure (one that violated natural parsing procedures), subjects would have to encode the figure many times and rotate it each time. Thus, the effects of angular disparity would be more pronounced for "bad" parts (and the difference in slopes would reflect the number of times the figure was re-parsed to find the "bad" part). If this account is correct, then this result says nothing about how entire figures are rotated or how the rotation procedure itself operates but merely

speaks to specific strategies subjects adopt when performing this particular task.

Another possibility is that these results are due to a visual comparison process, where subjects never rotate a mental image—even of a part. On this account, the results simply indicate that the detection task becomes increasingly difficult for "bad" parts when the subpattern is rotated further. A third possibility is that even if subjects did encode the entire figure and then mentally rotated it into congruence with the part, the results may be due to the image's becoming increasingly degraded with more rotation because it has had more time to fade. If so, "good" parts may still be relatively easily detectable at amounts of degradation that make it difficult to detect "bad" parts (see my discussion of Reed's, 1974, findings in chapter 7 of Kosslyn, 1980), resulting in detection times for "bad" parts that increase more sharply with increasing amounts of rotation. In short, then, one can draw no general conclusions about mental rotation from Pylyshyn's results.

Pylyshyn claims that attempts to modify his "holistic analogue view" bring one closer to the nonanalogue position. The issue is however, whether one can best explain the data without positing an analogue spatial medium. It simply is not clear why the contribution of this medium to the transformation stage is in any way detracted from by recognition that there are other stages used in performing any given task. There is nothing inherent in the position that the imagery representation system includes an analogue medium that even suggests that Pylyshyn's holistic analogue processing view (which includes many assumptions about process as well as structure) need be necessary or true.

The New Data

On Pylyshyn's mistaken view of our theory, scanning should always occur when images are inspected, and hence he takes his failure to find scanning in one situation as a disconfirmation of the theory. However, in point of fact our theory leads us to expect scanning to be required only if a to-be-classified part of the imaged object is depicted so far toward the periphery that it is too blurred to be easily categorized (recall that acuity decreases toward the periphery of the visual buffer; see Finke & Kosslyn, 1980; Kosslyn, 1978). For example, if one is focused on the tail of an imaged German Shepherd dog one will not be able to "see" the ears sharply enough to categorize their shape (round or pointed?) without scanning. (But even here scanning will be required only if the initial image is in fact used in performing the task, as I will explain shortly.)

Not only do Pylyshyn's results fail to disconfirm the theory, but the theory does not even have to strain to explain them. In fact, the results can be explained in numerous ways (and further experimentation is required to distinguish among these possibilities): One ready account of Pylyshyn's finding that subjects could judge relative orientation of imaged objects without scanning among them rests on the fact that one can "see" more than a single location in an image at the same time. Image inspection is not like viewing an object through a small hole in a piece of cardboard that must be moved around to infer a general shape. Thus, subjects conceivably could have performed the Pylyshyn task without having to scan, if they did in fact use an image. The size of the image is critical here, however, but it is impossible to estimate it, given the available description of the materials and procedure used in the new experiments. (Note that even if the stimulus configuration subtended too large a visual angle, the subjects may have formed their images at a smaller size unless carefully instructed not to do so—cf. Kosslyn et al., 1978.)

Another counter explanation of Pylyshyn's findings is as follows: In our theory we had to explain how people could scan from one location to another along a direct path. There were two basic options. First, we could have posited that people "see" the target position in the image and use this feedback to guide their scanning. Second, we could have posited that an abstract representation of position is encoded, allowing one to compute the direction to scan. The first alternative seems ruled out by our finding that people can scan a given distance to an overflowed part of an image as easily as to

a visible part of an image (see Kosslyn, 1978; Kosslyn et al., 1978). Further, when explaining how people generate images, we found it necessary to include relative location as an implicit part of the description of a part of an object. Thus, in our model each location on a map would be associated with a rough location specification. Hence, one can in fact perform Pylyshyn's task without recourse to an image, and Pylyshyn's results may simply indicate that his subjects did not in fact use imagery in performing the task. One way to distinguish between this notion and the first account offered above is to ask subjects to "zoom in" on the initial focus location, so that the rest of the image seems to overflow (as was done in the third experiment reported in Kosslyn, et al., 1978). If the task is no more difficult here, it would seem that subjects were not consulting the initial image in the condition where the entire array was visible at once.

Pylyshyn is disturbed by this last counterexplanation, which rests on use of an abstract form of spatial encoding. He worries that this undercuts the reasons for positing an analogue spatial medium in the first place. The point here is that we are dealing with a system that has many properties, some of which will be useful in performing some tasks and some of which will be useful in others. We did not simply introduce this abstract form of spatial representation to explain Pylyshyn's data, but needed it for entirely different reasons. And given that we did posit it, one can then ask under what conditions it will be useful.

Finally, what about the "blink" alternative explanation that Pylyshyn himself considered? We did not initially plan on building this property into our theory. However, as is discussed in the section on image transformations, we were stuck with the possibility of blink transformations, given our claims about image generation and the nature of the visual buffer (i.e., that it is a temporary store). Thus, we expect that subjects can image a transformed object in two ways: by shifting portions of it through the visual buffer or by letting the initial image fade and generating a new image of the object that is transformed in some way. If a blink transformation is used, the magnitude of the transform should not affect the time necessary to carry it out. And in fact, blink scans are equally easy over different distances (see chapter 8, Kosslyn, 1980). Pylyshyn argues that the amount of time required to generate images rules out a blink-transformation account of his data. This is an error. Pylyshyn used as his estimate of generation time the entire time subjects required to respond that they had generated an image after encoding a stimulus. This time includes encoding and response components of no interest here, since in a blink transform subjects already know which image to generate.

A better estimate of generation time is not the intercept but the slope, that is, the amount of time to generate an image of each additional part placed on an image. Interestingly, this time varies from about 50 to 150 msec for images of line drawings (see chapter 4, Kosslyn, 1980). Thus, even if time is required to adjust the values of the parameters used by the PICTURE process, and if the initial image requires some time to fade before the new one can be generated, the 300 additional msec in Pylyshyn's imagery condition may be enough to permit us to invoke a blink transform in explaining the data. In fact, times to perform blink scans and shift scans (where distance affects scan time) converge after relatively small distances are shift scanned (see chapter 8, Kosslyn, 1980). But this is really not to the point: Because different processing components are involved in performing the perceptual and imagery tasks, we cannot use the data from the perceptual task as a baseline for the imagery task. For example, eye movement control and execution add time in the perceptual condition but not the imagery one. Presumably, aspects of imagery processing add time in that condition relative to the perceptual one (e.g., the image may be less sharp than the percept and hence more difficult to inspect), but we have no way of knowing if the extra processing in the two tasks neatly cancelled each other out. (In fact, this seems very unlikely.) Thus, we simply cannot use the difference in times in Pylyshyn's two conditions to place constraints on the possible imagery processes. In addition, it is in general very dangerous to use estimates of

absolute times obtained in different experiments in the way Pylyshyn uses them, given the vagaries of different apparatuses, different subject populations, different experimenters, and so on. Finally, we cannot interpret the finding that distance affected time in the perceptual condition without knowing how well the locations were learned beforehand, how large the visual angle was at which the array was viewed, and the details of the instructions and procedure in general.

In short, then, Pylyshyn's new experiments were not so conceived that the predicted outcomes can serve to distinguish among the alternative positions.

The Tacit Knowledge Accounts

Pylyshyn (1979a, 1980, 1981) offers alternative accounts of the data from imagery experiments that rest on three major claims: First, the implicit task demands inherent in the instructions and the tasks themselves lead subjects to recreate as accurately as possible the perceptual events that would occur if they were actually observing the analogous situation. Second, subjects draw on their tacit knowledge of physics and the nature of the human perceptual system to decide how to behave in an experiment. Third, the subjects have the psychophysical skills necessary to produce the appropriate responses, for example by timing the interval between the onset of the stimulus and their pressing a key. There is one additional assumption that is critical for Pylyshyn's argument: The means by which tacit knowledge is invoked and used must not involve a spatial medium. That is, the issue at the heart of the differences in the alternative accounts centers on the existence of "depictive" representational structures in human memory, and thus a demonstration of the effects of tacit knowledge is relevant only if it can also be shown that this knowledge is not represented in a depictive form at some stage during the necessary processing. It would not be surprising if task demands and tacit knowledge sometimes affected imagery processing, given the long-standing claim that imagery can serve as a "dry run" simulation of the analogous physical events; if imagery were not able to be influenced by ideas, it would have only limited use in many forms of reasoning, such as those involved in the "thought experiments" reported by many distinguished scientists (see Shepard, 1978). Because Pylyshyn emphasizes the first two assumptions noted above in making his argument, it will behoove us to consider them in more detail.

Task Demands

Pylyshyn's argument depends on the notion that the instructions and the very nature of our tasks always led subjects to imagine what would happen in the analogous real event. However, this assumption fails to explain a number of our findings. Consider three examples: First, Kosslyn, Jolicoeur, and Fliegel (described in Kosslyn et al., 1979) performed a study in which subjects were asked to image an object and mentally "stare" at one end of it. Shortly thereafter, the name of a property was presented, and subjects were to judge the appropriateness of the property for the object as quickly as possible. It was stressed that the subject need not use the image in making the judgment. Interestingly, the distance from the point of focus to the property affected verification time only for properties that a separate group of subjects had rated to require imagery to verify (e.g., for a honeybee, "dark head"). No effects of distance were found for properties previously rated not to require imagery, even though the two kinds of items were randomly intermixed (and in fact the subjects had no idea that there were two different kinds of items). Pylyshyn's claim that some unspecified features of the task selectively evoke the "imagery habit" is merely an assertion of faith in the truth of his theory. This claim is tantamount to saying the results are explained by task demands because they must be explained by task demands—which is hardly a satisfactory account of the data.

Second, in another experiment, similar effects of the subjective size of parts of an image were obtained with first graders, fourth graders, and adults (see Kosslyn, 1976b). Further, in this experiment subjects began by evaluating a set of items without

being asked to use imagery. But after this verification task, the subjects were simply asked whether they had spontaneously tended to "look for" the named properties on images. When data from the first graders were analyzed in terms of which strategy was reported, I found effects of the size of properties only for those subjects who claimed to use imagery spontaneously. The result cannot be interpreted in terms of implicit demands in the instructions, since imagery was never mentioned at all.

Third, Finke and Kosslyn (1980) asked subjects to image pairs of dots moving toward the periphery and to indicate when the two dots were no longer distinct in the image. We also included a control group that was shown the stimuli and told the instructions we gave our experimental subjects. These people were asked to try to guess what our real subjects did and were explicitly asked not to use imagery in making these judgments. Interestingly, although these control subjects were able to guess that dots placed further apart would be "visible" at greater distances toward the periphery, they did not guess that distances increased less rapidly as dot separation increased. Further, the imagery field we measured was 1.83 times larger than that estimated by the control subjects. In general, the actual magnitudes estimated by the control subjects were consistently incorrect and were almost twice as variable as the corresponding data from the experimental subjects. Thus, although task demands may have evoked tacit knowledge about the general aspects of the imagery field, it did not make available all of the subtle properties that affect imagery processing. Finke (1980) presents numerous examples along these lines, all of which put strain on the kind of account advocated by Pylyshyn.

Tacit Knowledge of Perception

According to the tacit knowledge position, the imagery data are produced when subjects consider (without making use of analogue images) what something would look like if they were actually seeing it as it typically appears. I have three responses to this claim.

First, this position fails to provide accounts for the discoveries that imagery and perception share certain very counterintuitive properties, properties that people have never had the opportunity to discover in perception and that many people do not initially believe and find surprising when convinced. These properties are only manifest in highly novel laboratory settings, which precludes the subject from developing tacit knowledge about them from prior experience. Finke (1980) provides a good review of many of these studies in a recent issue of this journal, and I will not duplicate his efforts here. In addition to the studies he reviews, however, we have recently demonstrated that the geometrical properties of images affect processing in a way that is difficult to explain by appeal to tacit knowledge of perception or the physical world: Nancy Pennington and I (Note 1) asked people to image alternating black and white striped gratings and to mentally picture the gratings receding into the distance. We were interested in how far away the grating seemed at the point when the stripes blurred into a gray field. Interestingly, imaged vertical stripes seemed to blur at greater distances than oblique ones. This was true even if subjects began by seeing stripes at a given orientation and mentally rotated the stripes to a different orientation. None of our subjects were familiar with this "oblique effect" (which also occurs in perception and was found at the same magnitude in a separate group performing the perceptual analogue of the imagery experiment). In addition, the effects of different spatial frequencies (i.e., bar widths) were different in imagery and perception, which would not be expected if subjects were using tacit knowledge of perception to produce the expected results.

Second, the tacit knowledge view has considerable difficulty when properties of subjects' images differ from what they believe is typical about an object's appearance. For example, I found that the size (i.e., apparent angle subtended) at which images are spontaneously generated is different from that at which the objects are reportedly commonly seen (see Kosslyn, 1978). If subjects simply recall the typical perceived size of objects when asked to image them, this result

makes no sense. If different factors constrain imaged size (such as the extent of an analogue spatial medium within which images are formed) and typical viewed size, the result is not surprising.

Third, it is not enough simply to say that we cannot imagine some things because we could never see them (such as a four-dimensional cube). We should try to specify what it is about the perceptual system and the world that limits the range of possible percepts and images. One possibility, not ruled out by any of Pylyshyn's arguments, is that an analogue medium (such as the visual buffer we posit or the medium that supports a "2½-D sketch" in the Marr & Nishihara, 1978, model of perceptual processing) is used in both perception and imagery, and properties of this medium are one source of constraints on both what we can see and what we can imagine.

Evaluating the Theories

Theories are often evaluated by applying a set of abstract criteria, such as precision, generality, falsifiability, parsimony, and heuristic value. Let us consider the tacit knowledge theory and compare it to the Kosslyn and Shwartz theory on each of these criteria.

Precision. The tacit knowledge position has not been worked out in sufficient detail to evaluate its precision. The only claim is that there is some set of representations and processes that allows one to "simulate" physical or perceptual events without using analogue representations. In fact, Pylyshyn has presented some general metaprinciples for a theory, not a theory itself. With regard to the general issue of free parameters, it is worth noting that the tacit knowledge accounts may represent the ultimate in unspecified theories. These accounts seem virtually unconstrained and have no clearly specified parameters, free or otherwise.

The Kosslyn and Shwartz theory, in contrast, is precise enough to be implemented in a running computer simulation model. This general model produces specific models for numerous tasks (see the second half of Kosslyn, 1980), and these models in some cases account for quantitative relations in the data as well as the qualitative trends. Our theory is clearly more precise than Pylyshyn's.

Generality. To the extent that the tacit knowledge view rests on subjects' knowledge of or beliefs about how things typically happen in reality, the view has considerable generality. Our theory covers the domains of image generation, inspection, transformation, and the role of imagery in fact retrieval. Pylyshyn seems to intend his theory to cover almost all known imagery phenomena, and thus Pylyshyn's theory would appear the more general. But not all imagery phenomena simply mirror the analogous perceptual phenomena, and hence the tacit knowledge story can only go so far. That is, to the extent that imagery exhibits properties that differ from the analogous perceptual or physical ones, Pylyshyn's theory loses generality. My finding that images are not typically generated at the apparent size at which subjects believe the objects are usually seen (see Kosslyn, 1978) is a case in point, as is Pinker's (1980b) finding that the effects of 2- and 3-D distances are different in image scanning and actual perceptual scanning.

Falsifiability. Pylyshyn (1981) claims that tacit knowledge "could obviously depend on *anything* the subject might tacitly know or believe concerning what usually happens in the corresponding perceptual situations" (p. 34). Moreover, "the exact domain of knowledge being appealed to can vary from case to case" (p. 34). Thus, one can never be sure one has controlled for the effects of tacit knowledge in an experiment, in Pylyshyn's view. In other words, whereas the demand characteristics account may be disprovable, Pylyshyn's account, resting on implicit task demands, is sheltered from such a rude fate.

Our theory makes numerous predictions, and these predictions are subject to empirical test (see the second half of Kosslyn, 1980, and Shwartz, 1979). Because we eschew altering the properties of our theory merely to explain data, if predictions are not borne out, the theory can be falsified. In short, our theory is clearly more "vulnerable," more easily falsified, than Pylyshyn's.

Parsimony. Parsimony is not a property

that is inherent in a theory in and of itself (see Pinker, 1980a). Thus, one must compare two theories and decide which is the more parsimonious. Pylyshyn's theory appears more parsimonious than ours, but this is a consequence of his not filling in any details. Only when the tacit knowledge theory provides accounts at least as precise as ours can we then compare the two in terms of relative parsimony.

Heuristic value. The tacit knowledge account seems to have had little value thus far in producing new data (note that the results reported in Pylyshyn, 1981, are directed at undermining the analogue position, not supporting the tacit knowledge one per se). Further, mere demonstrations that tacit knowledge or task demands can affect image processing will neither demonstrate how they operate (which could involve manipulating images in a spatial medium) nor show that they usually underlie image processing. In contrast, the Kosslyn and Shwartz theory has proven to have a high degree of heuristic value. It has led us to collect data on a whole raft of imagery phenomena, data that all parties must now explain. The fact that Pylyshyn's theory has not led to important discoveries could simply reflect its relatively recent vintage, however, and thus it is premature to pass judgment on it with regard to this criterion.

Conclusions

Most of the criticisms of the Kosslyn and Shwartz theory were easily countered after the essentials of the theory were described, and Pylyshyn's recent alternative view did not prove incisive. Further, the data Pylyshyn introduced with an eye toward disproving an assumption of our theory were also found wanting. Not only was the prediction based on a confusion between the properties of structures and the properties of structure/process pairs, but the data themselves could be explained in multiple ways within the context of our imagery theory.

It is tempting to compare the present Pylyshyn critique with an earlier one (Pylyshyn, 1973) and to compare this reply with the one by Kosslyn and Pomerantz (1977). A number of points of similarity are apparent: First, both exchanges have been concerned with essentially the same issue (albeit approached in different ways), namely, whether image representations are different in kind from those underlying language processing. Second, both exchanges draw heavily on concepts drawn from outside psychology proper, from computer science, philosophy, and linguistics; the area of imagery representation now seems to have moved firmly into the domain of cognitive science.

More striking, however, are the dissimilarities between the present exchange and the earlier one: First, the earlier theorizing was at a much higher level of abstraction. In the present articles there is real movement toward development of explicit, detailed theories; we have proposed a detailed simulation model, and Pylyshyn is no longer merely asserting that processing is "propositional" but is offering some principles about how such processing would proceed. Second, in the earlier debate there were few data that bore directly on the nature of the imagery representation and processing system. We now have an abundance of data that place real constraints on theories (see Kosslyn & Shwartz, in press), and all parties are now concerned both with explaining the existing data and with collecting new data that speak to the issues. This is a very healthy sign, it seems to me. Third, there has been some real convergence in the alternative theories themselves. Both classes of accounts focus on imagery as a means whereby internal simulations can be conducted, and the main issue could be regarded as one concerning how far to carry this notion. Pylyshyn wants the spatial analogue medium itself to be merely simulated, whereas we want it to be an innately determined fixed structure that supports "depictive" data structures.

What, then, should we make of the apparent widespread dissatisfaction in the field that the so-called imagery–proposition debate has not yet been resolved? Some have even taken the view that the debate is in principle not capable of resolution (Anderson, 1978). Given the complexity of the issues involved, it seems overly optimistic to expect a speedy solution to such a knotty

problem. It is a mistake to think that theoretical issues in other fields have typically been resolved much more quickly. For example, it took about 30 years for physicists to resolve the issue of whether cathode rays were electromagnetic waves or trains of electrified particles (see D. L. Anderson, 1964), and it is not clear that we have any right to expect a speedier resolution to the present dispute. At the present juncture the best we can do is to continue to work out the empirical implications of the respective positions and to continue to collect new data that support one view while putting strain on others. Considerable effort and perseverance is required to carry this research to its conclusion, but we have good evidence that this effort will not be for naught—namely, that it has not been thus far. From any point of view, real progress has been forthcoming in a relatively short time.

Reference Note

1. Pennington, N., & Kosslyn, S. M. *The "oblique effect" in mental imagery.* Unpublished manuscript, Harvard University, 1980.

References

Anderson, D. L. *The discovery of the electron.* New York: Van Nostrand, 1964.

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

Anderson, J. R., & Bower, G. H. *Human associative memory.* New York: Winston, 1973.

Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. *Cognitive Psychology*, 1972, *3*, 472–517.

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

Finke, R. A. Levels of equivalence in imagery and perception. *Psychological Review*, 1980, *86*, 113–132.

Finke, R. A., & Kosslyn, S. M. Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 126–139.

Hayes-Roth, F. Distinguishing theories of representation: A critique of Anderson's "Arguments concerning mental imagery." *Psychological Review*, 1979, *86*, 376–392.

Hesse, M. B. *Models and analogies in science.* New York: Sheed & Ward, 1963.

Huttenlocher, J. Constructing spatial images: A strategy in reasoning. *Psychological Review*, 1968, *75*, 550–560.

Keenan, J. M., & Moore, R. E. Memory for images of concealed objects: A reexamination of Neisser and Kerr. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, *5*, 374–385.

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

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

Kosslyn, S. M. Using imagery to retrieve semantic information: A developmental study. *Child Development*, 1976, *47*, 434–444. (b)

Kosslyn, S. M. Measuring the visual angle of the mind's eye. *Cognitive Psychology*, 1978, *10*, 356–389.

Kosslyn, S. M. *Image and mind.* Cambridge, Mass.: Harvard University Press, 1980.

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

Kosslyn, S. M., & Jolicoeur, P. A theory-based approach to the study of individual differences in mental imagery. In R. E. Snow, P.-A. Federico, & W. E. Montague (Eds.), *Aptitude, learning, and instruction: Cognitive processes analysis of learning and problem solving* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1980.

Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. Category and continuum in mental comparisons. *Journal of Experimental Psychology: General*, 1977, *106*, 341–375.

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

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

Kosslyn, S. M., & Shwartz, S. P. A simulation of visual imagery. *Cognitive Science*, 1977, *1*, 265–295.

Kosslyn, S. M., & Shwartz, S. P. Visual images as spatial representations in active memory. In E. M. Riseman & A. R. Hanson (Eds.), *Computer vision systems.* New York: Academic Press, 1978.

Kosslyn, S. M., & Shwartz, S. P. Empirical constraints on theories of visual mental imagery. In A. Baddeley & J. Long (Eds.), *Attention and performance IX.* Hillsdale, N.J.: Erlbaum, in press.

Marr, D., & Nishihara, H. K. Visual information processing: Artificial intelligence and the sensorium of sight. *Technological Review*, 1978, *81*, 2–23.

Newell, A., & Simon, H. A. *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Pinker, S. Explanations in theories of language and imagery. *Behavioral and Brain Sciences*, 1980, *3*, 147–148. (a)

Pinker, S. Mental imagery and the third dimension. *Journal of Experimental Psychology: General*, 1980, *109*, 354–371. (b)

Pinker, S., & Kosslyn, S. M. The representation and manipulation of three-dimensional space in mental images. *Journal of Mental Imagery*, 1978, *2*, 69–84.

Pylyshyn, Z. W. What the mind's eye tells the mind's

brain: A critique of mental imagery. *Psychological Bulletin*, 1973, *80*, 1–24.

Pylyshyn, Z. W. Imagery theory: Not mysterious—just wrong. *Behavioral and Brain Sciences*, 1979, *2*, 561–563. (a)

Pylyshyn, Z. W. The rate of "mental rotation" of images: A test of a holistic analogue hypothesis. *Memory & Cognition*, 1979, *7*, 19–28. (b)

Pylyshyn, Z. W. Validating computational models: A critique of Anderson's indeterminacy of representation claim. *Psychological Review*, 1979, *86*, 383–394. (c)

Pylyshyn, Z. W. Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, 1980, *3*, 111–133.

Pylyshyn, Z. W. The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 1981, *87*, 16–45.

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

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

Shepard, R. N. The mental image. *American Psychologist*, 1978, *33*, 125–137.

Shwartz, S. P. *Studies of mental image rotation: Implications for a computer simulation of visual imagery*. Unpublished doctoral dissertation, Johns Hopkins University, 1979.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, *30*, 276–315.

Received May 5, 1980 ■