Psychological Review
1981, Vol. 88, No. 1, 16–45

0033-295X/81/8801-0016$00.75

The Imagery Debate: Analogue Media Versus Tacit Knowledge

Zenon W. Pylyshyn
Center for Advanced Study in the Behavioral Sciences
Stanford, California

The debate over the nature of mental imagery, especially with respect to the interpretation of recent findings on the transformation of images, has failed to focus on the crucial differences between the so-called "analogue" and "propositional" approaches. In this paper I attempt to clarify the disagreements by focusing on the alleged spatial nature of images and on recent findings concerned with "rotation" and "scanning" of mental images. It is argued that the main point of disagreement concerns whether certain aspects of the way in which images are transformed should be attributed to intrinsic knowledge-independent properties of the medium in which images are instantiated or the mechanisms by which they are processed, or whether images are typically transformed in certain ways because subjects take their task to be the simulation of the act of witnessing certain real events taking place and therefore use their tacit knowledge of the imaged situation to cause the transformation to proceed as they believe it would have proceeded in reality. The fundamental difference between these two modes of processing is examined, and certain general difficulties inherent in the analogue account are discussed. It is argued that the tacit knowledge account is more plausible, at least in the cases examined, because it is a more general account and also because certain empirical results demonstrate that both "mental scanning" and "mental rotation" transformations can be critically influenced by varying the instructions given to subjects and the precise form of the task used and that the form of the influence is explainable in terms of the semantic content of subjects' beliefs and goals—that is, that these operations are cognitively penetrable by subjects' beliefs and goals. Functions that are cognitively penetrable in this sense, it is argued, must be explained, at least in part, by reference to computational cognitive processes whose behavior is governed by goals, beliefs, and tacit knowledge rather than by properties of analogue mechanisms.

The study of mental imagery continues to be a major concern in cognitive psychology. Since regaining acceptance about 15 years ago, the study of processes underlying the sort of reasoning that is accompanied by perceptionlike experiences has become one of the focal points of the new mentalistic psychology. The purpose of this article is to comment on some of the recent theoretical work in this area in the light of the debate over the nature of mental imagery that has been recurring in the literature over the past 6 or 7 years. The various positions in this debate have been summarized in a number of places, including most recently in Shepard (1975, 1978), Kosslyn and Pomerantz (1977), Kosslyn, Pinker, Smith, and Shwartz (1979), Paivio (1977), Anderson (1978), and Pylyshyn (1973, 1978, 1979a, 1979b). What I shall do in this article is pick out what I consider to be the most substantive strand in this disagreement and discuss it in relation

This article was written while I was a fellow at the Center for Cognitive Science at MIT and at the Center for Advanced Study in the Behavioral Sciences at Stanford. The support of both these institutions is gratefully acknowledged, as well as the assistance provided by the Alfred P. Sloan Foundation and by a leave fellowship from the Social Science and Humanities Research Council of Canada. I would also like to thank Ned Block, Jerry Fodor, and Bob Moore for their helpful comments.

Requests for reprints should be sent to Zenon W. Pylyshyn, Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2.

to some of the most persuasive recent empirical findings and the most widely accepted theoretical accounts of these findings. For this purpose I shall make extensive reference to the overview article by Kosslyn et al. (1979), since it represents the most explicit formulation of the "imagistic" (or "pictorial" or "analogical") position to date. In doing this I shall be highly selective in the questions I shall address. There is much in the imagery literature that can be (and has been) debated, not all of which is equally significant from a theoretical standpoint. Thus one could argue over whether images are continuous or discrete, concrete or abstract, holistic or articulated, pictorial or discursive (whether they depict, like pictures, or refer, like descriptions), and whether they constitute a fundamentally different form of cognition or are merely a species of a single form used in all cognitive processing (and at what level they are considered to be the same or different). There have even been arguments over whether images are epiphenomenal or whether they are functional in cognition; but questions such as the latter cannot even be addressed until one takes a theoretical stand concerning the properties of images. One cannot say of something that it is or is not epiphenomenal until one has a clear statement of what that something is. For example, to the extent that *image* refers to what I experience when I imagine a scene, then surely *that* exists in the same sense that any other sensation or conscious content does (e.g., pains, tickles, etc.). If, on the other hand, *image* refers to a certain theoretical construct that is claimed to have certain properties (e.g., to be spatially extended) and to play a specified role in certain cognitive processes, then the appropriate question to ask is not whether the construct is epiphenomenal but whether the theoretical claims are warranted, and indeed whether they are true.

In my view, however, the central theoretical question in this controversy is whether the explanation of certain imagery phenomena requires that we postulate special types of processes or mechanisms, such as ones commonly referred to by the term *analogue.* I shall discuss one plausible interpretation of this notion—one that does indeed represent a fundamental difference in approach from the one I have been advocating. In addition to this issue, several of the other distinctions mentioned above can also be touched upon if we focus as sharply as possible on one particular claim often made in the imagery literature, namely, the alleged spatiality of images, and on a set of prototypical experimental findings that have been taken as establishing this particular property of images, namely, those that demonstrate "mental rotation" and "mental scanning" of images. In this regard I shall argue that the only real issue that divides the proponents of what has unfortunately become known as the "images versus propositions" debate is the question of whether certain aspects of cognition, generally (though not exclusively) associated with imagery, ought to be viewed as governed by tacit knowledge—that is, whether they should be explained in terms of processes which operate upon symbolic encodings of rules and other representations (such as beliefs and goals) or whether they should be viewed as intrinsic properties of certain representational media or of certain mechanisms that are not alterable in nomologically arbitrary ways by tacit knowledge. I have elsewhere referred to such mechanisms as constituting the "functional architecture" of the mind (Pylyshyn, 1980a, 1980b). In this article I will present arguments and evidence in support of the view that most of the empirical phenomena involving transformations of images (such as the image scanning results of Kosslyn et al., 1979) are better explained according to the tacit knowledge theory.

The Appeal to Properties of an Analogue Medium

In discussing the question of whether images are epiphenomenal, Kosslyn et al. (1979) assert that "none of the models of imagery based on Artificial Intelligence research treat the images that people report experiencing as functional representations" (p. 536). This strange yet widely held view is based in part on a misconception concerning what in fact is reported in imagery. As Hebb (1968) has pointed out, what people report is not properties of their image but of the objects that they are imaging. Such properties as color, shape, size, and so on are clearly properties of the objects that are being imagined. This distinction is crucial. The seemingly innocent scope slip that takes *image of object X with property P* to mean (*image of object X*) *with property P* instead of the correct *image of* (*object X with property P*) is probably the most ubiquitous and damaging conceptual confusion in the whole imagery literature.

To see that this slip is not a mere way of speaking but carries considerable weight in explanations of imagery phenomena, consider the case of the generally accepted "spatial" character of images. Take, for example, the elegant experiments by Kosslyn (1973, 1975) and by Kosslyn, Ball, and Reiser (1978) involving "mental scanning" of images, which show that the further away an item is from the place on an image that is currently being focused on, the longer it takes to see or focus on and report that item in the image. I shall take up the question of the interpretation of these results in the Tacit Knowledge and Mental Scanning section below. For the present I simply wish to point out that the story that goes with Kosslyn's interpretation inherits its plausibility and compellingness from a systematic equivocation over which particular entity has the property *length* in precisely the manner suggested in the previous paragraph.

For example there can be no disputing the Kosslyn et al. (1979) claim that "these results seem to indicate that images do represent metrical distance" (p. 537). But in the very next sentence this format-neutral claim becomes transformed into the substantive assertion that "images have spatial extent"—that is, that the image itself has rather than represents length or size. This transformation, moreover, is essential to the particular account of the scanning experiments that Kosslyn et al. wish to promote. That is because the naturalness of the scanning notion comes from the lawfulness of

$$T = \frac{D}{S}. \qquad (1)$$

In this equation, of course, T, D, and S are to be interpreted as real time, real physical distance, and real mean speed, respectively. If Equation 1 were literally applicable to the image, then this account of the scanning results would be a principled one, since the equation represents a universal principle or basic fact of nature. If, on the other hand, we were to keep with the first way Kosslyn et al. put their claim (viz., that images represent, rather than have distance), we would, instead, have to appeal to a different sort of regularity, one that might for instance be expressed roughly by

$$T = F(D', S'), \qquad (2)$$

where $D' = R_1(D)$ is some representation of distance using encoding R_1, $S' = R_2(S)$ is some representation of mean speed using encoding R_2, and F is a function that maps pairs of representations D' and S' onto real time such that for all distances, d, and speeds, s, it will be the case that $F[R_1(d),R_2(s)] = d \div s$.

Now Equation 2 is clearly not a law of nature. There can be no general universal law governing the amount of time that it takes to transform representations, since obviously that depends upon both the form of the representations and the available operations for transforming them. The equation $T = R_1(D) \div R_2(S)$ is far from expressing a nomological law. In fact, if F is to be realized computationally, we must view Equation 2 as asserting that there is some process, P, which, given the representations D' and S' as inputs (together with other specifications such as a beginning and ending state) takes T sec to complete, where T in this case has to equal $D \div S$. Obviously, unless the various representations and the process P are especially selected, Equation 2 will be false. For example, it is a nontrivial exercise to design an algorithm that always terminates after $D \div S$ sec when given two expressions representing the numerals for D and S (except for the degenerate case in which the algorithm calculates $D \div S$ and then simply waits idly for that amount of time to go by). Now by systematically leaving out the words *representation of* or by using ambiguous descriptions, such as saying that images "preserve relative metrical distances" (which can be interpreted as meaning either that they have or that they represent distances), it is possible to create the illusion of having the explanatory power provided by Equation 1 while at the same time avoiding the ontological claim that goes with it (viz., that images are actually laid out in space somewhere in the brain).

Another way to put the point about the relative explanatory power of the literal account based on Equation 1, compared with the representation account based on Equation 2, is in terms of the degrees of freedom in these two explanatory principles. If we assume that it is literally the case that physical space is involved, then the form of the relation among distance, speed, and time would be fixed as in Equation 1. If, on the other hand, only a representation of space is involved, and thus the regularity is expressed by Equation 2, the form of the function F is actually a free empirical parameter that is obtained by observing instances of the very phenomena that require explaining. This means that an explanation based on Equation 2 has more degrees of freedom and hence less explanatory power than an explanation based on Equation 1. Even more seriously, however, if we take Equation 2 as the appropriate formulation, then we need a theoretical account of why the relation holds and by what mechanisms it is realized. Even if we maintain that the cognitive system has evolved that way for one reason or another we still want to know what cognitive mechanisms are responsible for that behavior. There have traditionally been two approaches to providing such an account.

1.The first is to say that a subject makes Equation 2 come out (perhaps voluntarily, though often unconsciously) because he or she has tacit knowledge of Equation 1. In other words, regardless of the form of his or her representation, the subject knows that Equation 1 holds in the world and therefore makes it be the case (using some form of symbolic analysis, the exact nature of which need not concern us here) that the amount of time spent imagining the scanning will conform to this relation. We shall discuss this possibility in greater detail in the Tacit Knowledge and Mental Scanning section below.

2. The second way is to say that Equation 2 is the case because of properties of the representational medium. This is just to say that the observed function has that form as a consequence of the intrinsic lawful relations that hold among the particular physical properties that in fact represent distance and mean speed in the brain. For example, if distance were represented by the electrical potential between two points separated by a certain electrical capacitance, and mean speed were represented by current flow, then (within limits) the time taken would have the form given by Equation 2. This corresponds to what I would call the *analogue* view.

It should be appreciated that Alternatives 1 and 2 represent two fundamentally different ways of explaining the underlying process responsible for the observed behavioral regularities. Alternative 1 appeals to symbolically encoded facts about the world and to rules for transforming representations and drawing inferences. It is a "cognitivist" approach such as advocated by Fodor (1975, 1980), Chomsky (1980), Newell and Simon (1976), and others. On the other hand, Alternative 2 represents what I would call the analogue approach to mental representation and mental processing. The term *analogue* has been used to refer to a wide range of characteristics of models and representations covering everything from the mathematical continuity of representations to the simple requirement that the representation go through intermediate states representing the intermediate states that the actual system being represented would go through (e.g., Shepard, 1975). All of these capture something of what we intuitively mean by analogue. In my view, however, the only aspect of analogues that is relevant to the imagery debate (i.e., that differentiates among the major competing views) is the one raised by the distinction between Alternatives 1 and 2,—that is, an analogue process (represented by Alternative 2) is one whose behavior must be characterized in terms of intrinsic lawful relations among properties of a particular physical instantiation of a process, rather than in terms of rules and representations (or algorithms). Whenever people appeal to an "analogue representational medium" (e.g., Attneave, 1974) or to a "surface display" (e.g., Kosslyn, et al., 1979), they take it for granted that this medium incorporates a whole system of lawfully connected properties or intrinsic constraints (some of which have mathematical properties isomorphic to Equation 1 above) and that it is precisely this set of properties and relations that determines how objects represented in that medium will behave. Such people specifically contrast these accounts with ones like our Alternative 1, which claims that how the representation will behave is a function of what the person knows about the actual behavior of the things represented, rather than of properties of the medium in which it is represented.

Although there are various conceptions of what analogue processing is (as I suggested above), I suspect that the other senses are actually derivative from the sense I am adopting. Thus, for instance, any process can be made to go through an appropriate sequence of intermediate states, and even to do so in very small (quasi-continuous) steps—even a purely verbal process. Yet we would not want to count such a model as analogue if the mechanism were not naturally constrained to go through such a sequence. Thus we would count the process as analogue if its going through particular intermediate states were a necessary consequence of intrinsic properties of the mechanism or medium, rather than simply being a stipulated restriction that we arbitrarily imposed on a mechanism that could carry out the task in a quite different way. Palmer (1978) has taken a similar position with regard to the distinction between analogical and nonanalogical processes. From this, however, Palmer draws the unwarranted conclusion that only biological evidence will distinguish between the two forms of processing. But, as I have argued at some length (Pylyshyn, 1979b, 1980b), if we contrast analogue mechanisms with ones that operate on representations or tacit knowledge, the distinction can be seen to be a functional one that can be empirically decided by behavioral criteria. An example of one such criterion is discussed below. Other criteria are discussed in Pylyshyn (1979b, 1980b).

The Appeal to Tacit Knowledge

The distinction between analogue processes and rule-governed or cognitive processes (also referred to as computational or informational processes) is one that, in its most general form, needs to be drawn with some care, since after all, both are physically realized in the brain, although in quite different ways. The issue reduces to the question of when different forms of explanation of the behavior are appropriate. I have attempted to develop the general argument at length elsewhere (Pylyshyn, 1980a, 1980b). For the present purposes, however, a brief sketch of that discussion will do, since the only cases relevant to the imagery debate are unproblematic.

The operation of some processes can be explained perfectly well by giving an account of how various of their physical properties are causally connected, so that, for instance, altering some physical parameter here (e.g., by turning a knob) leads to specifiable changes in another parameter there because of some law connecting these two properties. Such a physical causal account will not do, however, to explain connections that are independent of the particular physical form the input takes yet that follow a single general principle that depends only on the semantic content of what might be called the input *message*. Thus, for example, if being told over the telephone that there is a fire in the building, seeing the word *fire* flash on a screen, hearing what you take to be a fire alarm, smelling smoke in the ventilator duct, seeing flames in the hallway, and so on without limit, all lead to the same building-evacuation behavior, the relevant generalization cannot be captured by a purely causal input–output story, since each such stimulus would involve a distinct causal chain and the set of such chains need have no physical laws in common. In that case, the generalization can only be stated by postulating internal belief and goal states (e.g., the belief that the building is on fire and knowledge about what one ought to do in such circumstances, as well as other tacit knowledge and the capacity to make inferences). Such processes are explainable only in terms of the mediation of rules and representations (since it is clear, for example, that neither a behavioral principle such as "Make sure you don't get too close to a fire" nor a logical principle such as modus ponens expresses a physical law and that they hold regardless of what kind of physical substance they are instantiated in).

A corollary of this explainability claim is that if a certain behavior pattern (or input–output function) can be altered in a way that is rationally connected with the meaning of certain inputs (i.e., what they refer to, as opposed to their physical properties alone), then the explanation of that function must appeal to operations upon symbolic representations such as beliefs and goals: It must, in other words, contain rule-governed cognitive or computational processes. A function that is alterable in this particular way is said to be *cognitively penetrable.* The criterion of cognitive penetrability (among other considerations) will be used in later discussions as a way of deciding whether particular empirically observed functions ought to be explained by Alternative 1 or by Alternative 2 above. Specifically, I shall maintain that if the form of certain image transformation functions reported in the literature can be altered in a particular sort of rationally explicable manner by changing what the subject believes the stimulus to be or by changing the subject's interpretation of the task (keeping all other conditions the same), then the explanation of the function must involve such constructs as beliefs, goals, or tacit knowledge, rather than the intrinsic properties of some medium—that is, some part of the explanation must take the form of Alternative 1.

The essence of the penetrability condition is this: Suppose subjects exhibit some behavior characterized by a function, f_1 (say, some relation between reaction time and distance or angle or perceived size of an imagined object), when they believe one thing, and some different function, f_2, when they believe another. Suppose further that which particular f they exhibit bears some logical or rational relation to the content of their belief: For example, they might believe that what they are imagining is very heavy and cannot accelerate rapidly under some particular applied force, and the observed f might then reflect slow movement of that object on their image. Such a logically coherent relation between the form of f and their belief (which we refer to as the "cognitive penetrability of f") must be explained somehow. Our claim is that to account for this sort of penetrability of the process, the explanation of f itself will have to contain processes that are rule governed or computational, such as processes of logical inference, and that make reference to semantically interpreted entities (i.e., symbols). The explanation cannot simply say that there are some causal (biological) laws which result in the observed function f (i.e., it cannot cite an analogue process), for exactly the same reason that an explanation of this kind would not be satisfactory in the building-evacuation example above: because the regularity in question depends on the semantic content (in this case of beliefs) and on logical relations that hold among these contents. Although in each particular case some physical process does cause the behavior, the general explanatory principle goes beyond the set of all observed cases (i.e., there may be token reduction but no type reduction of such principles to physical principles; (see Fodor, 1975). A process that is sensitive to the logical content of beliefs must itself contain at least some inferential (or other content-dependent) rule-governed process. It should be emphasized that cognitive penetrability refers not merely to any influence of cognitive factors on behavior but to a specific kind of semantically explicable (e.g., rational or logically coherent) relationship. The examples we shall encounter in connection with discussions of imagery will be clear cases of this sort of influence (for more on this particular point, see Pylyshyn, 1980a). It should also be noted that being cognitively penetrable does not prevent a process from having analogue components: It simply says that it should not be explained *solely* in terms of analogues with no reference to tacit knowledge, inference, or computational processes.

The concept of tacit knowledge—as a generalization and extension of the everyday notion of knowledge (much as the physicists' concept of energy is an extension of the everyday notion)—is one of the most powerful ideas to emerge from contemporary cognitive science (c.f. Fodor, 1968), although much remains to be worked out regarding the details of its form and function. It is already clear, however, that tacit knowledge cannot be freely accessed or updated by every cognitive process within the organism, nor can it enter freely into any logically valid inference. For example, much of it is not introspectable or verbally articulable (relevant examples of the latter would include our tacit knowledge of grammatical or logical rules, or even of most social conventions). A great deal needs to be learned about the control structures of the cognitive system that constrains our access to tacit knowledge in various elaborate ways. The existence of such constraints is no doubt what makes it possible for people to hold contradictory beliefs or to have beliefs that are only effective within certain relatively narrow classes of tasks. For example, it might well be that many people only have access to their tacit knowledge of physics when they are acting upon the world (e.g., playing baseball) or perhaps when they are engaged in something we call *visualizing* some physical process, but not when they have to reason verbally or answer certain kinds of questions in the abstract. Nonetheless, in all of these cases it would clearly be inappropriate to view such visualizing as being controlled by a medium or "surface display" that caused the laws of physics to hold in the image. A better way to view the cause of the regularities in the movement of objects in the visualized scene is in terms of subjects' tacit knowledge about the physical world and in terms of the inferences that they make from this knowledge. I shall consider other such examples in the next section when I argue that the appearance of autonomous unfolding of imagery sequences may be very misleading.

Incidentally, when one constructs a computer model using something called a matrix data structure rather than something called an analogue representational medium, one is not thereby relieved of the need to make a distinction between Alternatives 1 and 2. Because the existence of a computer model often carries the implication that there can no longer be any ambiguity or terminological confusions, it is worth examining one such model briefly.

In describing the Kosslyn and Shwartz (1977) computer model, Kosslyn et al. (1979) appear ready to admit that much of the model's explanatory and predictive capacity derives from what they refer to as the "cathode ray tube proto-model." In this proto-model there is no problem in seeing how a principled account of the scanning results can be derived. The CRT is a real physical device to which properties such as distance apply literally, and thus our earlier explanation of the observed scanning function applies in virtue of the applicability of Equation 1. But as we have already noted, such an explanation is a principled one only when it refers to a physical system whose intrinsic lawful behavior is described by Equation 1. Thus the explanatory power of the CRT proto-model only transfers to the human cognition case if there is something in the brain to which Equation 1 also applies. On the other hand, the version of the model that uses the matrix data structure, in which there is no actual physical CRT, lends itself equally to either one of the following two interpretations. In the first interpretation, the part of the model that contains the two-dimensional image (i.e., the 2-D matrix and its relevant access operations) is considered as merely a simulation of the physical screen, in which case the model really does assume the existence of a spatially laid out pattern in the brain. In the second interpretation, the matrix and the set of relevant accessing operations is viewed as a specific proposal for how the function F required by Equation 2 might be realized. In the latter case, however, when we give a theoretical interpretation of the claims associated with that part of the model, we still have a choice of the two basic views I have been calling Alternatives 1 and 2, exactly as we did when we were examining the informal account of the scanning results (e.g., Does the adjacency relation in the matrix represent subjects' knowledge that the elements referred to are next to one another, or is it an intrinsic constraint of that particular format?).

Appealing to a matrix in explaining certain imagery results is only useful if matrices constrain the representations or operations on representations in specified ways. If they do constrain the form of representations, then they function essentially as a simulation of an underlying analogue representational medium. (It might be noted in passing that if we were to take the matrix structure seriously, we would be stuck with the unavoidable conclusion that mentally represented space is necessarily nonisotropic. This is a formal consequence of the fact that a matrix is a tesselation of cells of some fixed shape and hence has certain essential nonisotropic properties. For example, if the cells are assumed to be square, then regardless of how fine we make them, scanning diagonally will be faster by a factor of the square root of two than scanning vertically or horizontally. Such an entailment cannot easily be glossed over, except by viewing the matrix as merely a metaphor for some unspecified spatial characteristics.)

The distinction between analogue and what I have sometimes referred to as propositional, but is perhaps better thought of as simply symbolic, is fundamental to a wide range of issues in the foundations of cognitive science (see Pylyshyn, 1980b). In the specific case of models of mental scanning, the distinction is important because Alternative 1 allows for the possibility that the results of mental scanning experiments may represent a discovery about what subjects believe and what they take the goal of the experiments to be, rather than a discovery about what the underlying mechanisms of image processing are. A consequence of the former alternative is that if subjects perceived the task differently or had different tacit beliefs about how the objects in question would move or about properties of space, then the experimental results could be quite different. On the other hand, if Alternative 2 were correct, then manipulation of such things as the form of the task and the instructions should not have a corresponding, rationally explainable effect (provided, of course, that imagery was still being used). Otherwise we would have to say that the medium changes its properties to correspond to what subjects believe about the world, in which case appealing to the existence of an analogue medium would serve no function.

Before turning to a discussion of some specific theoretical proposals, let me summarize the picture I have presented. Figure 1 illustrates the structure of alternatives available in explaining a variety of imagery findings. We can, first, choose a literal spatially extended brain-projection model. Although there is no a priori reason for excluding this alternative, it does raise some special problems if we try to explain the full range of imagery phenomena this way, and as far as I can tell, no one since Wertheimer has taken it seriously (though some, for instance Arbib, 1972, have come close). In any case the literal approach can be viewed as a special case of the analogue approach I shall be discussing in detail later. Continuing down our tree of alternatives, if we take the functional, as opposed to the literal or structural, approach, our task becomes to explain how this function could be realized by some possible mechanism (e.g., how Equation 2 could be realized in the case of mental scanning). Here we come to what I take to be the fundamental bifurcation between the two camps in the imagery debate, between those who advocate the analogue (or intrinsic property of a medium) view and those who advocate the symbolic or tacit knowledge view. Much confusion arises in this debate because there is considerable equivocation regarding exactly what the referents of ambiguous phrases such as *spatial representation* or *preserves metric spatial information* are intended to be from the point of view of this tree of alternatives. However, once the problem has been formulated so as to factor away the misleading implications associated with the use of a physical vocabulary, or a vocabulary that is appropriate for describing the represented domain as opposed to the psychological processes or mechanisms (see the next section), we are left with a basic empirical question: Which aspects of an organism's function are attributable to intrinsic (analogue) processes, and which are attributable to transactions on a knowledge base? It is to this empirical question that I now turn.

The Autonomy of the Imagery Process

It seems to me that the single most intriguing property of imagery, and the property that appears, at least on first impression, to distinguish it from other forms of deliberate rational thought, is that it has a certain intrinsic autonomy—both in terms of re-

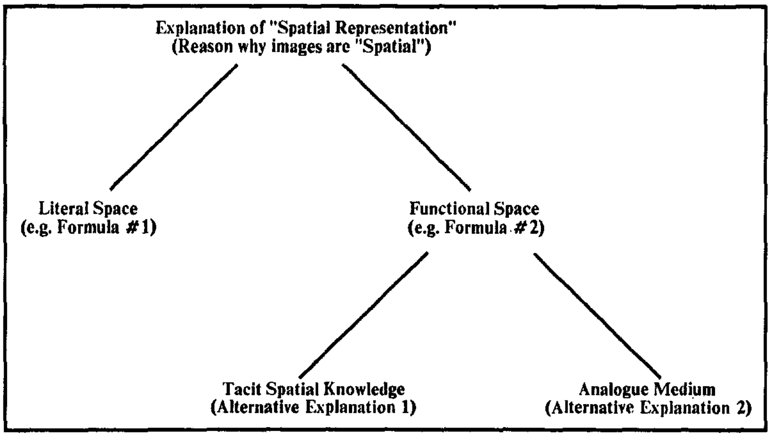

Figure 1. Theoretical positions on the nature of spatial representations.

quiring that certain properties of stimuli (e.g., shape, size) must always be represented in an image and with respect to the way in which dynamic imagery unfolds over time. Consider the second of these. The literature contains many anecdotes suggesting that in order to imagine a certain property, we first have to imagine something else (e.g., to imagine the color of someone's hair, we must first imagine the person's head or face; to imagine a certain room, we must first imagine entering it from a certain door; to imagine a figure in a certain orientation, we must first imagine it in a standard orientation and then imagine it rotating; to have a clear image of a tiny object, we must first imagine "zooming in" on it, and so on). Sometimes imagery even seems to resist our voluntary control. For example, in conducting a study of mental rotation of images, I instructed subjects to imagine moving around a figure, pictured as painted on the floor of a room. A number of subjects reported considerable difficulty in one of the conditions because the path of the imaginal movement was impeded by a wall visible in the photograph. They reported that they could not make themselves imagine moving around the figure because they kept bumping into the wall! Such responsiveness of the imagination to involuntary processes and unconcious control is one of the main reasons why imagery is associated with the creative process: It appears to have access to tacit knowledge and beliefs through other than deliberate intellectual routes.

Other examples involving imaginal movement may be even more compelling in this respect. Imagine dropping an object and watching it fall to the ground or throwing a ball and watching it bounce off a wall. Does it not naturally obey physical laws? Imagine rotating the letter *C* counterclockwise through 90°. Does it not suddenly appear to have the shape of a *U* without your having to deduce this? Imagine a square with a dot inside it. Now imagine the width of the square elongating until it becomes a wide rectangle. Is the dot not still inside the figure? Imagine the letters *A* through *E* written on a piece of paper in front of you. Can you not simply see by inspection that the letter *D* is to the right of the letter *B*? In none of these examples is there any awareness of what Haugeland (1978) calls "reasoning the problem through." The answer appears so directly and immediately available to inspection that it seems absurd to suggest, for example, that knowledge of topological properties of figures is relevant to the elongating square example or that tacit knowledge of the formal properties of the relation "to the right of" (that it is irreflexive, antisymmetric, transitive, connected, acyclic, etc.) is involved in the array of letters example. Such considerations have suggested to people that various intrinsic properties of imaginal representations are fixed by the underlying medium and that we exploit these fixed functional capacities when we reason imagistically. I believe that this intuition is the primary motivation for the widespread interest in analogue processes.

Now in general these are not implausible views. One should, however, be cautious in what one assumes to be an intrinsic function that is *instantiated* by the underlying biological structure, as opposed to one that is *computed* from tacit knowledge by the application of rules to symbolically represented beliefs, goals, and so on. In the previous section (as well as in Pylyshyn, 1980b) I have attempted to provide some necessary (though not sufficient) conditions for a function's being instantiated in this sense. The condition that I have found to be particularly useful in clarifying this distinction, especially in the case of deciding how to interpret observations such as those sketched above, is the one I called the *cognitive impenetrability criterion.* Recall that a function was said to be cognitively impenetrable if it could not be altered in a way that exhibits a coherent relation to the meaning of its inputs. For example, although a function might still count as being cognitively impenetrable if it varied with such things as practice or arousal level or ingestion of drugs, it would not be viewed as cognitively impenetrable if it changed in rationally explainable ways as a function of such things as whether a subject believes that the visually presented stimulus depicts a heavy object (and hence visualizes it as moving very slowly) or whether the subject views it as consisting of one or two figures, or as depicting an old woman or a young lady (in the well-known illusion), and as a consequence behaves in a way appropriate to that reading of the stimulus. I argued that cognitively penetrable phenomena such as the latter would have to be explained in terms of a cognitive rule-governed process, acting upon semantically interpreted representations and involving such activity as logical inferences, problem solving, guessing, associative recall, and so on, rather than in terms of the sort of natural laws that explain the behavior of analogue process.

Now many functions that appear at first to be biologically instantiated, and therefore alterable only in certain highly constrained law-governed respects, could turn out on closer inspection to be arbitrarily alterable in logically coherent ways by changes in subjects' beliefs and goals (i.e., they could turn out to be cognitively penetrable) and therefore to require a cognitive process account (based on appeal to tacit knowledge and rules). The tremendous flexibility of human cognition, especially in respect to the more central processes involved in thinking and commonsense reasoning, may very well not admit of many highly constrained (nonprogrammable) functions. It may illuminate the nature of the appeals to tacit knowledge if we consider some additional everyday examples. For instance, imagine holding in your two hands, and then simultaneously dropping, a large and a small object or two identically shaped objects of different weights. Which object in your image hits the ground first? Imagine turning a large heavy flywheel by hand. Now imagine applying the same torque to a small aluminum pulley. Which one completes one revolution in your image first? Imagine a transparent yellow filter and a transparent blue filter side by side. Now imagine slowly superimposing the two filters. What color do you see in your image through the superimposed filters? Form a clear and stable image of your favorite familiar scene. Can you now imagine it as a photographic negative, or as being out of focus, or in mirror image inversion, or upside down? Imagine a transparent plastic bag containing a colored fluid, being held open with four parallel rods at right angles to the mouth of the bag, and in such a way that the cross section of the bag is a square. Now imagine the four rods being moved apart so that, with the plastic bag still tight around them, the rods now give the bag a rectangular cross section. As you imagine this happening, does the fluid in the bag rise, fall, or stay at the same level (in other words, how does volume vary with changes of cross-sectional shape, perimeter remaining constant)? Imagine a glass half full of sugar and another nearly full of water. Imagine the sugar being poured into the glass with water in it. Examine your image to see the extent to which the resulting height of water rises (if at all).

These examples, it seems to me, are not in principle different from the ones in the first list I presented. For many people these imaginings also unfold naturally and effortlessly, without any need to reason through what would happen. Yet it seems clearer in these cases that whatever happens as the sequence unfolds under one's "mind's eye" is a function of what principles one believes govern the events in question. In fact, most people tend to get several of these examples wrong. Clearly the laws of dynamics or optics and the principles of geometry that determine the relation, say, between the perimeter and the area of a figure are not intrinsic (built in) to the representational media or to the functional mechanisms of the mind. Not only must one have tacit knowledge of them, but the way in which the imaginal events unfold naturally can usually be influenced with considerable freedom simply by informing the subject of the appropriate principle. Thus what seems to be a natural and autonomous unfolding process is cognitively penetrable—that is, it is under the control of an intellectual process, with all that this implies concerning the intervention of inferences and "reasoning through." As Harman (1973) has argued, our intuitions concerning when there are or are not inferences taking place must give way before the logical necessity to posit such processes. The mind, it seems, is faster than even the mind's eye.

Another particularly intriguing demonstration, by Ian Howard, shows that even in the case of a simple task involving the recognition of physically possible events, knowledge of physical principles is crucial—which suggests that in the case of imaging, such knowledge would be even more indispensable in explaining why images undergo transformations in certain systematic ways. Howard (1978) showed that over half the population of undergraduate subjects he tested could not correctly recognize trick photographs of tilted pitchers containing colored fluids whose surface orientations were artificially set at various anomalous angles relative to the horizontal. Using the method of random presentation and repeating the study with both stereoscopic photographs and motion pictures, Howard found that the subjects who failed the recognition task, for levels as much as 30° off horizontal, could nevertheless correctly report that the fluid surface was not parallel to shelves visible in the background, thus showing that it was not a failure of perceptual discrimination. What was particularly noteworthy, however, was that post-experimental interviews, scored blindly by two independent judges, revealed that every subject who scored perfect on the recognition test (i.e., no stimulus with orientation more than 5° off horizontal failed to be correctly classified as anomalous) could clearly articulate the principle of fluid level invariance, whereas no subject who made errors gave even a hint of understanding the relevant principle. In this case, unlike what typically happens in other areas such as phonology, evidence of the relevant knowledge was obtainable through direct interviews. Even when such direct evidence is not available, however, indirect behavioral evidence that tacit knowledge is involved can frequently be obtained—for example by demonstrating cognitive penetrability of the phenomenon to new knowledge.

Once examples of this kind are presented, no one finds the claim that tacit knowledge is required for some imaginings at all surprising. In fact, Kosslyn et al. (1979) admit that both image formation and image transformation can be cognitively penetrable and hence not explainable by appealing to properties of the imaginal medium. But given that these examples are not distinguishable from the earlier ones in terms of the apparent autonomy of their progression, why do we continue to find it so compelling to view phenomena such as those associated with mental scanning as providing evidence for an intrinsic property of the imaginal medium (or, as Kosslyn et al. put it, for the "spatial structure of the surface display")? Is it because processes such as mental scanning are more resistant to voluntary control? Is it really inconceivable that we could "search for" objects in a mental image without passing through intermediate points, or that we could compare two shapes in different orientations without necessarily first imagining one of them being at each of a large number of intermediate orientations?

I believe that what makes certain spatial operations resistant to a natural interpretation in terms of knowledge-governed reasoning through is primarily the "objective pull" I discussed in Pylyshyn (1978), which results in the tendency to view the cognitive process in terms of properties of the represented objects (i.e., the semantics of the representation) instead of the structure of the representation itself (i.e., the syntax of the representation). This tendency, however, leads to a way of stating the principles by which mental processes operate that deprives the principles of any explanatory value. This involves appealing to principles that are expressed in terms of properties of the represented object rather than in terms of the structure or form of the representation itself. But expressing a principle in terms of properties of the represented domain begs the question of why processing occurs this way. The mechanism has no access to the properties of the represented domain *except insofar as they are encoded in the form of the representation itself.* Consequently, a principle of mental processing must be stated in terms of the formal structural properties of its representations, not in terms of what they are taken to represent in the theory.

Consider the following example. In describing their model, Kosslyn et al. (1979) take care to abstract the general principles of operation of the model—a step that is essential if the model is to be explanatory. Such principles include, for example, "Mental images are transformed in small steps, so the images pass through intermediate stages of transformation" (p. 542), or "The degree of distortion will be proportional to the size of the transformational step" (p. 542). Shepard (1975) also cites such principles in discussing his image transformation results.

In each case these principles only make sense if terms such as *size* or *small steps* refer to the *represented* domain. In other words, the intended interpretation of the first of the above principles would have to be something like the following: Representations are transformed in such a way that successive representations correspond to small differences in the scene being depicted. However, what we need is a statement that refers only to the structure of the representation and its accessing process. We need to be able to say something like, "Representations are transformed in small structural steps or degrees," where *small* is relative to a metric defined over the formal structure of the representation and process. For example, relative to a binary representation of numbers and a machine with a bit-shifting operation, the formal (or syntactic) transformation from a representation of the number 137 to a representation of the number 274 is smaller than the transformation from a representation of 137 to a representation of 140 (since the former requires only one shift operation), even though it clearly corresponds to a larger (semantic) transformation in the represented domain of abstract numbers. It is thus important to distinguish between the intrinsic syntactic domain and the extrinsic semantic domain when speaking (typically ambiguously) about transformations of representations. Not only are the two domains logically distinct, but, as we have seen, they involve two quite different similarity metrics and their behavior is governed by quite different principles.

Cognitive principles such as those invoked by Kosslyn et al. (1979), Shepard (1975), and Anderson (1978) would only be theoretically substantive (i.e., explanatory) if they specified (a) how it was possible to have formal operations that had the desired semantic generalization as their consequence—that is, how one could arrange a formal representation and operations upon it so that small steps in the formal representations corresponded to small steps in the represented domain—and (b) why these particular operations, rather than some other ones that could also accomplish the task, should be used (this is the issue of making the underlying theory principled, or restricting its degrees of freedom by reducing the number of free parameters in it). Simply asserting that representations do, as a matter of fact, have this property (because, for example, they are said to "depict" rather than merely "represent"; c.f. Kosslyn et al., 1979) is not enough. One reason that it is not enough is that such a property is simply stipulated in order to conform to the data at hand: It is a free empirical parameter. Another reason is that, as with our earlier cases, there are two distinct options available to account for how this can happen. They correspond to our options 1 and 2: We can appeal either to tacit knowledge or to intrinsic properties of a representational medium. In other words, we can make the account principled by relating the process either to the rationality of the method adopted, given the organism's goals and tacit knowledge, or to the causal relations that hold among physical properties of the representational medium.

Before examining in greater detail the proposal that many imagery phenomena, including specifically those dealing with mental scanning, should be explained in terms of tacit knowledge, we need to touch on two additional points, since they frequently muddy the discussion. For this reason we shall make a brief digression.

Constraints of Habit and the Executive Retreat

The first point is that we must distinguish here, as we do in other areas of theory evaluation, between what typically or frequently or habitually happens and what *must* happen because of some lawful regularity. Insofar as the origin of visual imagery no doubt lies in visual experience, and insofar as we frequently see things happen in certain ways, this could easily influence the way in which we typically imagine certain kinds of events. For example, since our visual experiences are primarily with common, middle-sized objects, certain typical speeds of acceleration, deceleration, and trajectory shapes are much more common than others. If this sort of experience did influence our tendency to most frequently imagine movements in certain ways, it would clearly not be something attributable to the nature of the mechanism or to properties of the representational medium.

Examples of habitual modes of processing determined by the nature of our experience rather than by our fixed functional capacities are frequent in many areas of cognition. For example, when we learn to read, the teacher monitors our performance auditorily. Consequently, we first learn to read out loud and later to suppress the actual sound. As a result, many of us continue to read by converting written text into a phonetic form prior to further analysis. But there is good reason to believe that this stage of processing is not a necessary one (e.g., Forster, 1976). In fact, there is a plausible view that this habitual mode of processing is responsible for the slow reading speed that some of us suffer from and that we can be trained to abandon.

Although knowing the habitual modes of processing is useful if one is interested in describing the typical, or in accounting for variance, or in developing practical tools (say for education), providing an explanation of behavior requires that we understand the nature of the underlying mechanism (or medium). This, in turn, requires that we empirically establish how imaginal processing is constrained, or which of its functions are independent of particular beliefs or goals (i.e., we must discover the cognitively impenetrable properties of imaginal processing). Thus it becomes important to ask which particular characteristics (if any) the use of imagery forces on us, rather than to report the strategies that tend to go along with the use of imagery. For example, rather than asking whether, in using imagery, subjects *typically* take more time to locate (or otherwise focus their attention on) objects represented as more distant or to report the presence of features in an image that they describe as smaller, we should ask whether subjects *must* do so whenever the image mode of representation is being used. A variety of experimental findings (e.g., Kosslyn et al., 1979, Bannon, Note 1; Spoehr & Williams, Note 2) have demonstrated that, left to their own devices, people habitually solve certain kinds of problems (typically ones involving metrical or geometrical properties) by visualizing some physically possible event taking place (e.g., they imagine themselves witnessing the stimulus changing in certain characteristic ways). Yet no one, to my knowledge, has tried to set up an experimental situation in which subjects were discouraged from carrying out the task by this habitual means to determine whether they were constrained to do so by some cognitive mechanism or medium. Later in this paper I shall report several studies carried out with this goal in mind.

The second point that needs clearing up concerns the grain of truth in Anderson's (1978) claim that the form of representation cannot be determined unequivocally by appeal to behavioral data alone. Anderson's argument is that for any model using some particular form of representation, one can always conjure up another behaviorally indistinguishable model that uses a different form of representation simply by making compensatory changes in the accessing process. As I tried to show (Pylyshyn, 1979b), this cannot be done in general without considerable loss in explanatory power. One of the considerations that led Anderson to the indeterminism view is the apparent unresolvability of the so-called imaginal versus propositional representation debate. No sooner does one side produce what they take to be a damaging result than the other side finds a way to compensate for this by adjusting the process that accesses the representation while leaving fixed the assumed properties of the representation itself. In my view, the lesson to be learned from this observation is simply that in adjudicating such a debate, as in determining the correct interpretation of any empirical phenomenon, one should not appeal solely to data (for even adding neurophysiological and any other class of data does not solve the problem, since no finite amount of data alone can ever uniquely determine a theory), but one should also consider the explanatory power of the model—that is, how well it captures important generalizations, how constrained it is (i.e., how many free parameters it has), how general it is, and so on. Anderson's views on this criterion notwithstanding, the issue is not fraught with vagueness and subjectivity. Although it clearly is not simple to apply in practice, the notion of explanatory power is crucial to the conduct of scientific inquiry, inasmuch as we need to distinguish between such predictive devices as curve fitting or statistical extrapolations and genuine cases of law-like explanatory principles. I shall have more to say about the issue of predictive versus explanatory adequacy in the concluding section.

These remarks are intended as an introduction to the most common rejoinder made to arguments (such as those based on the informal examples considered above, as well as experimental observations of cognitive penetrability) that we should attribute the properties and behavior of images to tacit knowledge rather than to intrinsic properties of an imaginal medium. This rejoinder consists of the counterproposal that we retain the analogue medium but simply modify the processes that generate, transform, and interpret the representation and thus enable the analogical model to account for such findings. For example, my objections to certain particular analogue models of image processing, which are based on demonstrations that certain imaginal processes are cognitively penetrable, can often be sidestepped by merely adding an additional layer of executive process that varies the generation and transformation of images in response to cognitive factors. Sometimes this sort of executive overlay can be made to produce the desired behavior in an imagery model, but rarely can this be done without adding various ad hoc contrivances and consequently losing explanatory power.

Consider, for the sake of a concrete example, the case of the phenomenon called "mental rotation." I claimed (Pylyshyn, 1979a) that the operation of mentally rotating a whole image is not one of the functions that is instantiated by the knowledge-independent functional capacities of the brain and hence should not be explained by appealing to properties of some analogue medium. My conclusion in this case was based on the empirical finding that the slope of the relationship between the relative orientations of two figures and the time it takes to carry out certain comparisons between them (such as deciding whether they are identical)—which is generally taken as the behavioral measure of rate of mental rotation—depends on various cognitive factors such as figural complexity and the difficulty of the actual postrotation comparison task. In these studies the difficulty of the comparison task was varied by requiring subjects to decide whether one figure was embedded in the other and then varying the "goodness" or gestalt value of the embedding (see Pylyshyn, 1979a, for details). One counterargument to my conclusion suggested by Kosslyn et al. (1979)—and one, incidentally, that I considered in my original paper—is that these findings are compatible with a holistic analogue view because an executive process might have determined, on the basis of some property of the stimulus or probe figure, what rate of rotation to use and set this as the value of a parameter to the rotation function. A number of responses might be made to this suggestion.

First, although nothing in principle prevents one from making rotation rate a parameter of the analogue, such a proposal weakens the explanatory power of the model considerably, for any behavioral property, not just rate of rotation, could be made a parameter (including, for example, the form of the function relating reaction time to orientation). The more such parameters there are, the more the model becomes an exercise in curve fitting. Because these parameters are not constrained a priori, each contributes to the degrees of freedom available for fitting the observed data and hence detracts from the explanatory power of the model. This is another way of saying that unless we have some independent means of theoretically assigning a rotation rate to each stimulus, such a parametric feature of the model will be completely ad hoc. Thus there is considerable incentive to try to account for the comparison times in some principled way, either on the basis of some intrinsic property of the representational medium or else in terms of some aggregate characteristic of the cognitive process itself (such as, for example, the number of basic operations carried out in each condition).

It is this very consideration that leads me to agree with Kosslyn et al. (1979) when, in discussing the relative merits of the analogue view of mental rotation—as opposed to the alternative propositional account proposed by Anderson (1978), in which a parameter describing the orientation of a figure is incrimentally recomputed—they state, "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 in the other?" (p. 545). From this point of view, the analogue proposal is clearly less ad hoc, since it posits a universal constraint that is associated with the medium itself (not with that phenomenon alone) and that is therefore not a free empirical parameter. However, it should be noted that this account is only principled when it refers to the intrinsic analogue medium model of rotation, not to the symbol-structure (or matrix) view I discussed earlier. In the latter case the principles (e.g., rotation proceeds by application of transformations that correspond to small distances) appeal to properties of the represented domain, rather than to intrinsic properties of the representation, precisely as does Anderson's ad hoc incrimental parameter adjustment proposal. In both cases no principle based on some independently determined property of the representation or of the structure of the process is given for why this should be so. On the other hand, the trouble with the principled analogue view is that it appears to be false as it stands, as I argued in Pylyshyn (1979a).

Despite the fact that there is strong incentive to account for observed properties, such as rotation rate, in the principled ways suggested above, it could still turn out that the best we can do at the present time is to appeal to something like a rotation rate parameter. In fact, as I have argued (Pylyshyn, 1980b), there will necessarily be some primitive functions that are themselves not explainable in terms of symbol manipulation processes. These constitute what I called the "functional architecture" of the mind. There is no a priori reason why a one-argument version of the operator ROTATE(speed) cannot be such a process. In such a model, the speed parameter would be viewed as being adjusted by some physical means (e.g., a digital-to-analogue converter) on the basis of a cognitive analysis of the stimulus and the subject's beliefs and goals. The question of whether this is the correct story is ultimately an empirical one, just as was the original question of whether ROTATE is an instantiated analogue function (i.e., an instrinsic property of the medium, or what I have called Alternative 2). What has to be done in that case is to expose this new proposal to empirical tests such as those that assess cognitive penetrability. Of course, as I pointed out, each such retreat from the original holistic analogue hypothesis brings us closer to Alternative 1 (the tacit knowledge explanation), as more of the determinants of the phenomena are put into the class of logical analyses and inferences.

Finally it should be pointed out that the particular proposal for parametrizing rotation rate does not, in any case, apply to the experimental results I reported. In these studies the slope of the reaction time versus angle curve was shown to be a function not merely of properties of the stimulus figure or the comparison figure but of the difficulty of the comparison task itself (i.e., the task of deciding whether the probe was an embedded subfigure of the stimulus). The holistic analogue model assumes that the comparison phase can only be carried out after the stimulus figure has been rotated into the appropriate (independently determined) orientation (indeed, that is the very phase of the process that is responsible for the linear relation between angle and time). In fact Kosslyn et al. (1979) appear to implicitly accept this particular order of events when they propose the alternative that "people may choose in advance slower rates for 'worse' probes" (p. 546). The trouble with this alternative, however, is that we found rotation rate to be not only a function of the nature of the stimulus and of the probe but also of the relation between them, specifically of how well the probe fits as an embedded part of the stimulus. Since this particular feature of the comparison phase cannot be known in advance of rotation, it could not possibly be used as a basis for setting a rate parameter.

Now I have no doubt that one could come up with some kind of executive process that utilized a holistic analogue and yet exhibited different rates of apparent rotation for the different conditions, as observed. However the fact that one could design such an executive would itself be of little interest. It would require some strong independent motivation for going to such lengths in order to retain the analogue rotation components. As I suggested earlier, the main attraction of the analogue model is that it is both principled (i.e., it posits a universal property of mind) and constrained. It is constrained because it permits only one way to transform an image of a figure in one orientation into an image of that figure in another orientation, in contrast with the unlimited number of ways in which an arbitrary symbol structure can in principle be transformed. This constraint would have constituted a powerful explanatory principle. But now as we locate more and more of the explanatory burden in the executive process, there remains less and less reason to retain the ROTATE analogue operation, although as I stated above, we will always need to posit some knowledge-independent functional properties or capacities (i.e., *analogue* in my sense).

Having thus outlined two general methodological considerations that need to be kept in mind when interpreting empirical findings bearing on the contrast between the intrinsic property of the medium view and the tacit knowledge view, I am ready to consider the specific case of the mental scanning phenomena in some detail.

Tacit Knowledge and Mental Scanning

In examining what takes place in studies such as those discussed by Kosslyn et al. (1979), it is critical to note the difference between the following two tasks:

1a. Solve a particular problem by using a certain prescribed form of representation, or a certain medium or mechanism.

1b. Attempt to recreate as accurately as possible the sequence of perceptual events that would occur if you were actually observing a certain real event happening.

The reason this difference is critical is that quite different criteria of success apply in these two cases. For example, solving a problem by using a certain representational format does not entail that various incidental properties of a known situation even be considered, let alone simulated. On the other hand, this is precisely what is required of someone solving Task 1b. In this case failure to duplicate such conditions as the speed with which an event occurs would constitute a failure to carry out that task correctly. Take the case of imagining. The task of imagining that something is the case, or of considering an imagined situation in order to answer questions about it, does not entail (as part of the specification of the task itself) that it take any particular length of time. On the other hand, the task of imagining that an event is actually happening before your very eyes does entail, for a successful realization of this task, that you consider as many as possible of the characteristics of the event, even if they are irrelevant to the discrimination task itself, and that you attempt to place them into the correct time relationships.

For instance, in discussing how he imaged his music, Mozart claimed (see Mozart's letter reproduced in Ghiselin, 1952), "Nor do I hear in my imagination, the parts *successively*, but I hear them, as it were, all at once" (p. 45). He felt that he could hear a whole symphony in his imagination all at once and apprehend its structure and beauty. Clearly he had in mind a task that is best described in terms of 1a. Even the word *hear*, taken in the sense of having an auditorylike imaginal experience, need not entail anything about the duration of that experience. We can be reasonably sure that Mozart did not intend the sense of *imagining* implied by 1b, simply because if what he claimed to be doing was imagining witnessing the real event of, say, sitting in the Odeon Conservatoire in Munich and hearing his Symphony Number 40 in G Minor being played with impeccable precision by the resident orchestra under the veteran Kapellmeister, and if he had been imagining that it was actually happening before him in real time and in complete detail—including the minutest flourishes of the horns and the trills of the flute and oboe, all in the correct temporal relations and durations—then he would have taken very close to 22 minutes for this task. If he had not taken that long to imagine it, this would only signify that he had not quite been doing what he had alleged, that is, he had not been imagining witnessing the actual real event in which every note was being played at its proper duration, or else we might conclude that what he had in fact been imagining was not a good performance of his symphony. In other words, if it takes *n* sec to witness a certain event, then an accurate mental simulation of the act of witnessing that same event should also take *n* sec, simply because how well the latter task is performed is by definition dependent on how accurately it mimics various properties of the former task. On the other hand, the same need not apply merely to the act of imagining that the event *has* a certain set of properties, that is, imagining a situation to be the case but without the added requirements as specified in the 1b version of the task. These are not empirical assertions about how people imagine and think: They are simply claims about the existence of two distinct natural interpretations of the specification of a certain task.

Applying this to the particular case of mental scanning, one must be careful to distinguish between the following two tasks that subjects might set themselves:

2a. Using a mental image and focusing your attention on a certain object in that image, decide as quickly as possible whether a second named object is present elsewhere in that image.

2b. Imagine yourself in a certain real situation in which you are viewing a certain scene and are focusing directly on some particular object in that scene. Now imagine that you are looking for (or scanning toward, or glancing up at, or seeing a speck moving across the scene toward, etc.) a second named object in the scene. When you succeed in imagining yourself finding (and seeing) the object (or when you see the speck arrive at the object), press this button.

The relevant differences between Tasks 2a and 2b should be obvious. As in the previous examples, the criteria of successful completion of the task are different in the two cases. In particular, Task 2b includes, as part of its specification, such requirements as that subjects should attempt to imagine various intermediate states (corresponding to ones that they believe would be passed through in actually carrying out the corresponding real task) and that they spend more time visualizing those episodes that they believe (or infer) would take more time in the corresponding real task. The latter conditions are clearly not part of the specification of Task 2a, as there is nothing about Task 2a that requires that such incidental features of the visual task be considered in answering the question. In the words of Newell and Simon (1972), the two tasks have quite different "task demands."

To show that subjects are actually carrying out Task 2b in the various studies reported by Kosslyn (and therefore that the proper explanation of these findings should appeal to subjects' tacit knowledge of the depicted situation rather than to properties of their imaginal medium), I shall attempt to establish several independent points. First, it is independently plausible that the methods used in experiments reported in the literature should be inviting subjects to carry out Task 2b rather than Task 2a. Second, the arguments against experimental demand effects raised by Kosslyn et al. (1979) do not bear on the above proposal. Third, this alternative view has considerable generality and can account for a variety of imaginal phenomena. And fourth, there is independent experimental evidence showing that subjects can indeed be led to carry out Task 2a rather than Task 2b, and when they do, the increase in reaction time with increase in imagined distance disappears.

Task Demands of Scanning Experiments

With respect to the first point, all published studies that I am aware of in which larger image distances led to longer reaction times used instructions that quite explicitly required subjects to imagine witnessing the occurrence of a real physical event. In most scanning experiments subjects are asked to imagine a spot moving from one point to another, although in a few (e.g., in Kosslyn, 1973; Kosslyn, Ball, & Reiser, 1978, Experiment 4) they were asked to imagine shifting their attention or their glance from one imagined object to another in the same imagined scene. In each case, what subjects were required to imagine was a real physical event (since terms like *move* and *shift* refer to physical processes) about the duration of which they would clearly have some reasonable tacit knowledge. For example, they would know implicitly that it takes a moving object longer to move through a greater distance, that it takes longer to shift one's attention through greater distances (both transversely and in depth), and so on. Although subjects may or may not be able to state these regularities, they plainly do have that tacit knowledge, as evidenced by the critical precision necessary to make realistic motion pictures by splicing pan and zoom sequences. (The exact time relationships needed to make such sequences appear realistic, especially in the case of splicing together takes of slower and more deliberate movements of actors and of points of view, seem to depend on one's prior interpretation of the actions. Hence the process involved in detecting poor film editing, like the process of imagining realistic scenarios, would seem to be knowledge dependent and therefore cognitively penetrable.)

The Arguments Against Demand Characteristics

Kosslyn et al. (1979) appear to recognize some of the force of the tacit knowledge position, but in responding to it they concern themselves only with the possibility that "experimental demand characteristics," or unintentional influences due to the experimental setting and subjects' expectations, might have been responsible for the outcome of the experiments. Although recent results by Richman, Mitchell, and Reznick (1979) and Mitchell and Richman (1980) indicate that phenomena such as those found in mental scanning experiments can be brought about by experimental demand factors, it has not yet been established that this is in fact the correct explanation for all such results. Kosslyn et al. have argued that it is unlikely that demand factors could explain all their results. On the other hand, neither have they provided any definitive control studies to rule out this alternative (the "pseudo-experiment" described by Kosslyn et al. is inadequate in this respect, inasmuch as simply asking subjects what they expect is the best way to invite acquiescence effects, as opposed to genuine expectations or other types of demand biases).

However, whether the case for experimental demand effects will stand up to empirical tests or whether the Kosslyn et al. counterarguments are correct is not relevant to the present proposal. There is a major difference between the contaminating effects of experimental demands, or subjects' expectations of the outcome or their desire to please, and the entirely legitimate task demands, or requirements placed on the solution process by the specifications of the task itself. In the latter case what is at issue is not a contamination of results but simply a case of subjects solving the task as they interpret it (or as they choose to interpret it, for one reason or another) by bringing to bear everything that they know about a class of physical events, which they take to be the ones that they are to imagine witnessing. If they take the task to be the one characterized in Task 2b, then they will naturally attempt to reproduce a temporal sequence of representations corresponding to the sequence they believe would arise from actually viewing the event of scanning across a scene (or seeing a spot move across the scene). Thus, beginning with the representation corresponding to "imagining seeing the initial point of focus," the process would continue until a representation was arrived at which corresponded to "imagining seeing the named point." Of course, according to this way of viewing what is going on, there is no need to assume that the process halts as a *result* of a certain imagined state's being reached, or when a certain visual predicate is satisfied. It could just as plausibly stop when some independent psychophysical mechanism had generated a time interval corresponding to an estimate of expected duration (we know such mechanisms exist, since subjects can generate time intervals corresponding to known magnitides with even greater reliability than they can estimate them; c.f. Fraisse, 1963). In other words, it could just as easily be independently estimated time intervals that drive the imagined state changes.

For the purpose of this account of the scanning results, we need assume little or nothing about intrinsic constraints on the process or even about the content of the sequence of representations that are generated. Such a sequence could, for example, simply consist of a sequence of beliefs such as that the spot is *now here* and *now* it is *there*—where the locative demonstratives are pointers into the symbolic representation being constructed and updated. Though the sequence is almost certainly more complex than this, there is no need to assume that it is constrained by any special property of the representational medium, as opposed to simply being governed by what subjects believe or infer about some likely intermediate stages of the event being imagined and about the relative times at which they would occur. Now such beliefs and inferences could obviously depend on anything that the subject might tacitly know or believe concerning what usually happens in the corresponding perceptual situations. Thus the sequence could in one case depend on tacit knowledge of the dynamics of physical objects, in another on tacit knowledge of some aspects of eye movements or of what happens when one has to glance up or refocus on a more distant object, or even on tacit knowledge of how long it takes to notice or to recognize certain kinds of visual patterns (e.g., it might even take subjects longer to imagine trying to see something in dim light or against a camouflage background for this reason). Thus none of the examples and contrary evidence that Kosslyn et al. (1979) cite against one or another of the alternative "experimental demand" explanations is to the point here, since the exact domain of knowledge being appealed to can vary from case to case, as is to be expected if imagining is viewed as a species of commonsense reasoning, as opposed to a process that has access to a special sort of representational medium with extraordinary functional properties (e.g., being characterized by Euclidean axioms).

Sometimes experiments involving superimposing images on actual visual stimuli have been cited against the demand characteristics view (e.g., Kosslyn et al., 1979). However, such experiments differ from studies of imaginal thinking in several important respects that make them largely irrelevant to the present discussion. When a subject is instructed to view a display and then to imagine a stationary or a moving pattern superimposed on it (as in the studies by Hays, 1973; Finke, 1979; Shulman, Remington, & McLean, 1979; and those mentioned in Shepard, 1978), there is no need to posit an internal medium of representation to explain the stable, *geometrical* relationships that hold among features of the resulting construction. The perceived background itself is all we need in this case. For example, when a subject thinks of an imaginary spot as being *here* and then *there* (as in the discussion above), the locative terms can in this case be bound to places in a perceptual construction that are under direct stimulus control and that are generally veridical with respect to relative spatial locations. This is essentially equivalent to binding the internal symbols to the actual places in the stimulus, which, being in the actual stimulus, will maintain their locations relative to one another regardless of subjects' beliefs about space or about what they are viewing (assuming only that perception is free from major time-varying distortions). In fact, Pylyshyn, Elcock, Marmor, and Sander (1978) have developed a model of how indexical binding of internal symbols to primitive perceptual features can be carried out within a limited-resource computational system and how such bindings can be used by the motor system to enable it to, say, point to the bound features.

Thus in such superposition cases if, for instance, the subject imagines a spot moving from perceived location A to perceived location B, then all that is required to ensure that the spot crosses some location C is (a) that the successive locations where the point is imagined to be actually correspond to a certain path on the stimulus (i.e., that the successive mental locatives in fact refer to a certain sequence of adjacent places on the stimulus) and (b) that place C actually be on that path, somewhere between A and B. In the pure imagery case, by contrast, the corresponding notions of *path, lying on,* and *between* are not available in the same literal sense (i.e., there are only representations of paths). In other words, subjects must not only imagine the spot to be moving with respect to an imagined background but they must have tacit knowledge of such things as that if C lies between A and B, then going from A to B requires passing through C. Another way to put this is to say that the geometrical properties of the layout that is being viewed (e.g., the relative locations of features in it) remain fixed because of the way the world being viewed is (in this case, rigid), and different geometrical characteristics of the layout can simply be "noticed" or "perceived" by the viewer, including the relative position of a place being attended to (i.e., a place that is bound to an internal locative indexical symbol). On the other hand, what remains fixed and what can be noticed in a purely constructed image depends either on intrinsic properties of some medium of representation or on subjects' tacit knowledge about the behavior of the sorts of things they are imagining and their ability to draw inferences from such knowledge—exactly the dichotomy we are examining. This issue is closely connected with the general problem of reasoning about actions, which in artificial intelligence research raises a technical problem called the "frame problem." The relevance of such issues to the imagery controversy is discussed in Pylyshyn (1978, 1980b).

The Generality of the Tacit Knowledge View

With respect to the generality of explanations based on appeal to tacit knowledge, one could point to a variety of findings that fall nicely within this explanatory framework. For instance, the list of illustrative examples presented in the last section shows clearly that in order to imagine the episode of seeing certain physical events, one needs to have access to tacit knowledge about physical regularities. In some of these cases one might even say that one needed an implicit theory, since a variety of related generalizations must be brought to bear in order to correctly predict what some imagined process would do (e.g., the sugar solution or the color filter case). In other cases simply the knowledge (or recollection) that certain things typically happen in certain ways and that they take certain relative amounts of time will suffice.

Several of Kosslyn's findings, allegedly revealing properties of the "mind's eye," might also be explainable on this basis—including the finding (Kosslyn, 1975) that it takes longer to report properties of objects when the objects are imagined as being small. Consider that the usual way to inspect an object is to take up a viewing position at some convenient distance from the object which depends on its size (and in certain cases on other things as well; e.g., consider imagining a deadly snake or a raging fire). So long as we have a reasonably good idea of the object's true size we would imagine viewing it at the appropriate distance. Now if someone instructed me to imagine some object as especially small, I might perhaps think of myself as being further away or as seeing it through, say, the wrong end of a telescope. In any case if I were then asked to do something, such as report some of its properties, and if the instructions were to imagine that I could *actually see* the property I was reporting (which was the case in the experiments reported), or even if I simply chose to make that my task for some obscure reason, I would naturally try to imagine the occurrence of some real sequence of events in which I went from seeing the object as small to seeing it as big enough so I could easily discern certain details (i.e., I would very likely take the instructions as meaning that I should carry out Task 1b). In that case I would probably imagine something that was in fact a plausible visual event, such as a zooming-in sequence (and indeed this is what many of Kosslyn's subjects reported). If that were the case then we would naturally expect the time relations to be as actually observed.

Although the above story may sound quite a bit like the one Kosslyn (1975) himself gives, there is one difference that is crucial from a theoretical standpoint. In this version of the account, no appeals need to be made to knowledge-independent functional properties of a medium, and especially to properties of a *geometrical* sort. The representational medium, although it no doubt has some relevant intrinsic properties that restrict how things can be represented, plays no role in accounting for any of the particular phenomena we have been examining. These phenomena are seen as arising from (a) subjects' tacit knowledge of how things typically happen in reality and (b) their ability to carry out such psychophysical tasks as to generate time intervals corresponding to inferred durations of certain possible physical events. This is not to deny the importance of different forms of representation, of the nature of such inferential capacities as alluded to above, or of the nature of the underlying mechanisms. It is simply to suggest that the particular findings we have been discussing do not necessarily tell us anything about such matters.

Although we intuitively feel that the visual image modality (or format, or medium) severely constrains both the form and the content of potential representations, it is no easy matter to say exactly what these constraints are (and the informal examples given earlier should cast at least some suspicions on the validity of such intuitions in general). It seems clear, for example, that we cannot image any arbitrary object whose properties we can describe, and this does give credence to the view that images are more constrained than descriptions. Although it is doubtlessly true that imagery is in some sense not as flexible as discursive symbol systems (such as language), it is crucial to know the nature of this constraint before we can say whether it is a constraint imposed by the medium or merely a habitual way of doing things or of interpreting the task demands, or whether it might even be a limitation attributable to the absence of certain knowledge or a failure to draw certain inferences. Once again I would argue that we cannot say a priori whether certain constraints implicated in the use of imagery ought to be attributed to the functional character of the biological medium of representation (the analogue view) or to the subject's possession and use (either voluntarily or habitually) of certain tacit knowledge.

Consider the following proposals made by Kosslyn et al. (1979) concerning the nature of the constraints on imagery. The authors clearly take such constraints to be given by the intrinsic nature of the representational medium. They suggest that something they call the "surface display" (a reference to their cathode ray tube proto-model) gives imagery certain fixed characteristics. For example, they state,

> 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 display has. Thus we predict that a person cannot at will make his surface display four-dimensional, or non-Euclidean. (p. 549)

Now it does seem to be obviously true that one cannot image a four-dimensional or non-Euclidean space. Yet the very oddness of the supposition that we might be able to do so should make us suspicious as to the reason for this.

To see why little can be concluded from this fact, consider the following. Suppose a subject insisted that he or she could imagine a non-Eulidean space. Suppose further that mental scanning experiments were consistent with this claim (e.g., scan time conformed to, say, a city block metric). Would we believe this subject or would we conclude that what the subject really did was to

> *simulate* such properties in imagery by filling in the surface display with patterns of a certain sort in the same way that projections of non-Euclidean surfaces can be depicted on two-dimensional Euclidean paper? (Kosslyn et al., 1979 p. 547)

Of course we would conclude the latter. But the reason for doing so is exactly the reason we gave earlier for discounting one possible interpretation of what Mozart might have meant when he claimed to be able to imagine a whole symphony instantaneously. That reason, you will recall, had to do entirely with the implications of one particular sense of the phrase *imagine a symphony*—namely that the Task 2b sense demands that certain conditions be fulfilled. If we transpose this to the case of the spatial property of visual imagery, we can see that this is also the reason why the notion of imagining four-dimensional space in the sense of Task 2b is incoherent. The point is sufficiently central that it merits a brief elaboration.

Let us first distinguish, as I have been insisting we should, the sense of imagining (call it *imagine$_t$ X*) that means to *think of X* or to consider the hypothetical situation that X is the case (or mentally construct a symbolic model or a mental description of a possible world in which X is the case) from the sense of imagining (call this one *imagine$_s$ X*) that means to imagine that you are *seeing X* or to imagine yourself observing the actual event X happening. Then the reason for the inadmissibility of four-dimensional or non-Euclidean imaginal space becomes clear, as does its irrelevance to the question of what the properties of an imaginal medium are. The reason we cannot imagine$_s$ such spaces is that they are not the sorts of things that could be seen. Our inability to imagine$_s$ such things has nothing to do with intrinsic properties of a surface display, but with a lack of a certain sort of knowledge: We do not know what it would be like to see such a thing. We have no idea, for example, what kind of configuration of light and dark contours there would have to be, what sorts of visual features would need to appear, and so on. Presumably congenitally color-blind people cannot imagine$_s$ a colored scene for similar reasons. In this case it would hardly seem appropriate to attribute this failure to something's being wrong with their surface display. On the other hand, we do know, in nonvisual (i.e., nonoptical) terms, what a non-Euclidean space is like, and we can imagine$_t$ there being such a space in reality (certainly Einstein did) and thus solve problems about it. Perhaps, given sufficient familiarity with the facts of such spaces, we could even produce mental scanning results in conformity with non-Euclidean geometries. There have frequently been reports of people who claimed to have an intuitive grasp of four-dimensional space in the sense that they could do such things as mentally rotate a four-dimensional tesseract and imagine$_s$ its three-dimensional projection from a new four-dimensional orientation (for example, Hinton, 1906, has an interesting discussion of what is involved). If this were true, then they might be able to do a four-dimensional version of the Shepard mental rotation task.

Of course if we drop all this talk about the geometry of the display and consider the general point regarding the common conceptual constraints imposed on vision and imagery, there can be no argument: Something is responsible for the way we cognize the world. Whatever it is probably also explains both the way we see it and the way we image it. But that is as far as we can go. From this we can no more draw conclusions about the geometry, topology, or other structural property of a representational medium than we can draw conclusions about the structure of a language by considering the structure of things that can be described in that language. There is no reason to believe that the relation is anything but conventional—which is precisely what the formalist (or computational) version of functionalism claims (see Fodor, 1980).

Incidentally, the distinction between the two senses of imagine discussed above also clarifies why various empirical findings involving imagery might tend to occur together. For example, there is a brief report in the authors' response section of Kosslyn et al. (1979) of a study by Kosslyn, Jolicoeur, and Fliegel showing that when stimuli are sorted according to whether subjects tend to visualize them in reporting certain of their properties (i.e., whether subjects typically imagine$_s$ them in such tasks), then it is only those stimulus–property pairs that are classified as mental image evokers that yield the characteristic reaction time functions in mental scanning experiments. But that is hardly surprising, since anything that leads certain stimuli to be habitually processed in the imagine$_s$ mode will tend to exhibit all sorts of other characteristics associated with imagine$_s$ processing—including the scanning time results and such phenomena as the "visual angle of the mind's eye" or the relation between latency and imagined size of objects (see the summary in Kosslyn et al., 1979). Of course nobody knows which features of a stimulus or task tend to elicit the imagine$_s$ habit or why some stimuli should do so more than others, but that is not a problem that distinguishes the analogue from the tacit knowledge views.

Some Empirical Evidence

Finally I shall consider some provisional evidence suggesting that subjects can be induced to use their visual image to carry out a task such as 2a that does not entail imagining oneself seeing a natural sequence of events happening. Recall that the question was whether mental scanning effects (i.e., the linear relation between time and distance) should be viewed as evidence for an intrinsic property of a representational medium or as evidence for such things as what tacit knowledge (of geometry and dynamics) people have and what they take the task to be. If the former were the correct interpretation, then it must not merely be the case that people usually take more time for retrieving information about more distant objects in an imagined scene. That could arise, as we have already noted, merely from some habitual or preferred way of imagining or a preferred interpretation of the task demands. If the phenomenon is due to an intrinsic property of the imaginal medium, then it must be a necessary consequence of using this medium; that is, the linear (or at least monotonic) relation between time and represented distance must hold whenever information is being accessed through the medium of imagery.

As it happens, there exists a strong preference for interpreting tasks involving doing something imaginally as tasks of type 1b—that is, as requiring one to imagine$_s$ an actual physically realizable event happening over time. In most of the mental scanning cases, it is the event of moving one's attention from place to place or of witnessing something moving between two points. It could also involve imagining such episodes as drawing or extrapolating a line and watching its progression (which may be what was involved, for example, in the Spoehr & Williams study, Note 2). But the question remains: Must a subject imagine such a physically realizable event in order to access information from an image, or more precisely, in order to produce an answer which the subject claims is based on examining the image?

A number of studies have been carried out in our laboratory which suggest that conditions can be set up so that a subject uses an image to access information, yet does so without having to imagine the occurrence of some particular real life temporal event (i.e., the subject can be induced to imagine$_t$ rather than imagine$_s$). I will mention only two of these studies for purposes of illustration. The design of the experiments follows very closely that of experiments reported in Kosslyn, Ball, and Reiser (1978; see Bannon, Note 1, for more details). Subjects had to memorize a map containing approximately seven visually distinct places (e.g., a church, a castle, a beach) up to the criterion of being able to reproduce it with the relative location of places within 6 mm of the correct location. Then they were asked to image the map in front of them and to focus their attention on a particular named place, while keeping the rest of the map in view in their mind's eye. We then investigated various conditions in which they were given different instructions for what to do next, all of which (a) emphasized that the task was to be carried out exclusively by consulting their image and (b) required them to notice, on cue, a second named place on the map and to make some discriminatory response with respect to that place as quickly and as accurately as possible.

So far this description of the method is compatible with the Kosslyn et al. (1978) experiments. Indeed, when we instructed subjects to imagine a speck moving from the place of initial focus to the second named place, we obtained the same kind of strongly linear relation between distance and reaction time as did Kosslyn et al. When, however, the instructions specified merely that subjects should give the compass bearing of the second place—that is, to say whether the second place was N, NE, E, SE, and so forth of the first, there was *no* relation between distance and reaction time. (In this experiment subjects were first given practice in the use of the compass direction responses and were instructed to be as fast and accurate as possible within the resolution of the eight available categories. In postexperiment interviews, subjects reported that they carried out the task by consulting their image, as they had been instructed.)

This result suggests that it is possible to arrange a situation in which subjects use their images to retrieve information and yet do not feel compelled to imagine the occurrence of an event that would be described as scanning their attention between the two points (i.e., to imagine$_s$). Although this result was suggestive, it lacked controls for a number of alternative explanations. In particular, since a subject must in any case know the bearing of a second place on the map before scanning to it (even in Kosslyn's experiments), one might wish to claim, for independent reasons, that in this experiment the relative bearing of pairs of points on the map was retrieved from a symbolic, as opposed to imaginal, representation, in spite of subjects' insistence that they did use their image in making their judgements. Although this tends to weaken the imagery story somewhat, since it allows a crucial spatial property to be represented off the display (and so raises the question, Why not represent other spatial properties this way?) and because it discounts subjects' reports of how they were carrying out the task in this case while accepting such reports in other comparable situations, it is nonetheless one possible avenue of retreat.

Consequently, a second instructional condition was investigated, aimed at making it more plausible that subjects had to consult their image in order to make the response, and to make it more compelling that they must have been focused on the second place and mentally seeing both the original and the second place at the time of the response. The only change in the instructions that was made for this purpose was to explicitly require subjects to focus on the second place after they heard its name (e.g., *church*) and, using it as the origin, give the orientation of the first place (the place initially focused on) relative to the second. Thus the instructions strongly emphasized the necessity of focusing on the second place and of actually seeing both places before making the orientation judgment. Subjects were not told how to get to the second place from the first, but only to keep the image before their "mind's eye" and to use this image to read off the correct answer. In addition, for reasons to be mentioned shortly, the identical experiment was run (using a different group of subjects) entirely in the visual modality, so instead of having to image the map, subjects could actually examine the map in front of them. Eight subjects were run in the image condition and eight in the vision one. Each subject was given 84 trials, thus providing four times for each of the 21 interpoint distances.

What we found was that in the visual condition, there was a significant correlation between response time (measured from the presentation of the name of the second place) and the distance between places, whereas no such relation held in the imaginal condition. In doing the analysis, distances were grouped into small, medium, and large, and a linear regression was carried out on the grouped data. In the visual condition there was a significant correlation between distance and reaction time ($r = .50$, $p < .05$). In the imaginal condition there was no significant correlation ($r = -.03$, *ns*). The mean reaction time in the visual condition was 2.60 sec and in the imaginal condition was 2.90 sec. Such results indicate quite clearly that even though the linear relation between distance and time (the scanning phenomenon) is a frequent concomitant of imaging a transition between seeing two places on an image, it is not a necessary consequence of using the visual imagery modality and consequently that it is not due to an intrinsic (hence knowledge- and goal-independent) functional property of the representational medium for visual images.

Yet perhaps not surprisingly, results such as these can be accommodated without too much trouble by the Kosslyn et al. model. That model has been conveniently provided with the option of "blinking" its way to a second location—or of regenerating a new image from symbolic information. In that case it would clearly be able to respond in fixed time, regardless of the distance between places. Several remarks can be made concerning this alternative.

First, the existence of both scan and blink transforms can be used simply to ensure that no empirical data could falsify the assumption of an intrinsic medium of representation. Whether or not this is the case depends on what, if any, additional contraints are placed on the use of these transforms. Kosslyn et al. (1979) do suggest that people will use whichever transform is most efficient. Thus they ought to scan through short distances but blink over longer ones. This, however, presupposes that they know in advance how far away they will have to move over the image—and hence that distance information is available for arbitrary pairs of places without consulting the image and without requiring scanning. Clearly this assumption is inconsistent with the original assumption regarding how spatial information is accessed from images. We shall return to this point briefly in the concluding section, when we consider where the predictive power of such imagery models comes from.

Second, if the correct explanation for our results is that subjects used the blink transformation and hence generated new images instead of using their initial ones to locate the second place (as they were instructed to do, and as they reported having done), then we should be able to see the effect of this in the overall response times. Since it took our subjects 1 or 2 sec to generate the initial image, it is very unlikely that they were regenerating a completely new image and making the required orientation judgment in the 2.9 sec it took for them to respond. Perhaps they were only regenerating the two critical places within the existing outline in their image. But even that seems implausible for the following reason. The average reaction time to make orientation judgments in the visual condition, where no image had to be generated, was only 300 msec shorter than the average time to make the judgment in the imagery condition. This indicates that if an image had to be regenerated in the imagery condition, as assumed by the blink transformation explanation, it would have taken less than 300 msec to regenerate such an image. Since, according to Kosslyn, Reiser, Farrah, and Fliegel (Note 3), it usually takes several seconds to generate even simple images—and never less than 1 sec even for images containing only one simple part—there is insufficient time to both regenerate parts of an image and make an orientation judgment in the total 2.9 sec it took subjects to respond. Hence subjects could not have been using a blink transformation to regenerate their image in that case.

These experiments demonstrate that, at least in the one situation investigated, images can be examined without the putative constraints of the surface display postulated by Kosslyn and others. It is also reasonable to expect that other systematic relations between reaction time and image properties may disappear when appropriate instructions are given that are designed to encourage subjects to interpret the task as in 1a instead of 1b. For example, if subjects could be induced to generate what they considered small but highly detailed and clear images, then the effect of image size on time to report the presence of features (e.g., Kosslyn, 1975) might disappear as well. There is even some evidence that this might be the case from one of Kosslyn's own studies. In one of the studies reported in Kosslyn et al. (Note 3), the time to retrieve information from images was found to be independent of the size of the image. From the description of this experiment, it seems that a critical difference between it and the earlier ones (Kosslyn, 1975), in which an effect of image size was found, is that in this case subjects had time to study the actual objects, with instructions to practice generating equally clear images of each of them, and were also tested with these same instructions (which

I assume encouraged them to entertain equally detailed images at all sizes). Thus it seems that it is possible, when subjects are encouraged to have detailed information readily available, for subjects to put as fine a grain of detail as they wish into their imaginal constructions (though presumably the total amount of information in the image is still limited along some dimension, even if not the dimension of resolution). Unlike the case of real vision, however, such imaginal vision need not be limited by problems of grain or resolution or any other difficulty associated with making visual discriminations. Of course, as we have already noted, subjects can exhibit some of the behavioral characteristics associated with such limitations (e.g., taking longer to recall fine details), but that may very well be because they know what real vision is like and are simulating the relevant behavior as best they can, rather than because of the intrinsic nature of the imaginal medium.

Conclusions: What Is the Theoretical Claim?

It has often been said that imagery models (such as that of Kosslyn & Shwartz, 1977, or Shepard, 1975) contribute to scientific progress because they make correct predictions and because they motivate further research. Although I would not want to deny this claim, it is important to ask what it is about such imagery models that carries the predictive force. It is my view that there is only one empirical hypothesis responsible for the predictive success of the whole range of imagistic models and that nearly everything else about such models consists of free empirical parameters added ad hoc to accommodate particular experimental results. The one empirical hypothesis is just this: *When people imagine a scene or an event, what goes on in their minds is in many ways similar to what goes on when they observe the corresponding event actually happening.*

It is to the credit of both Shepard (1978) and Paivio (1977) that they recognize the central contribution of the perceptual metaphor. For example Shepard (1978) states,

> Most basically, what I am arguing for here is the notion that the internal process that represents the transformation of an external object, just as much as the internal process that represents the object itself, is in large part the same whether the transformation, or the object, is merely imagined or actually perceived. (p. 135)

Paivio (1977) has been even more direct in recognizing and approving of the metaphorical nature of this class of models when he asserts,

> The criteria for a psychological model should be what the mind can do, so why not begin with a psychological metaphor in which we try to extend our present knowledge about perception and behavior to the inner world of memory and thought. . . . The perceptual metaphor . . . holds the mirror up to nature and makes human competence itself the model of mind. (p. 71)

One difficulty with metaphorical explanation in general is that by leaving open the question of what the similarities are between the primary and secondary objects of the metaphor, it remains flexible enough to encompass most eventualities. Of course this open-endedness is also what gives metaphors their heuristic and motivational value and is what provides the feeling of having captured a system of regularities. But in the case of the perceptual metaphor for imagery, this sort of capturing of regularities is, to a large extent, illusory, because it is parasitic upon our informal commonsense knowledge of psychology and our tacit knowledge of the natural world. For example, I have argued that the reason I imagine things happening more or less the way that they actually do happen in the world is not because my brain or my cognitive endowments are structured to somehow correspond to nature but simply because I know how things generally happen—because I have been told, or have induced, what some of the general principles are. In other words I have a tacit physical theory which is good enough to predict most ordinary everyday natural events correctly most of the time. Now the claim that our imagery unfolds the same way as our perceptual process trades on this tacit knowledge in an even more insidious way, because it does so in the name of scientific explanation.

The story goes like this. The claim that imagery is (in some ways) like perception has predictive value because it enables us to predict that, say, it will take longer to mentally scan longer distances, to report the visual characteristics of smaller imagined objects, to rotate images through larger angles, to mentally compare more similar images, and so on. It does this because we know that these generalizations hold in the corresponding visual cases. But notice that the reason we can make such predictions is not that we have a corresponding theory of the visual cases. It is simply that our tacit commonsense knowledge is sufficiently accurate to provide us with the correct expectations in such cases. An accurate analogy would be to give, as a theory of Mary's behavior, the statement that Mary is a lot like Susan, whom we know very well. This would enable us to perhaps make very accurate predictions of Mary's behavior, but it would scarcely qualify as an adequate explanation of why she behaves as she does. Another parallel, even closer in spirit to the metaphorical explanation of imagery, would be if we gave as the explanation of why it takes longer to rotate a real object through a greater angle that this is the way we typically perceive it happen, or if we explained why it takes more time to visually compare two objects of similar size than two objects of very different sizes by saying that this is what happens in the mental comparison case. In both these cases we would be able to make the correct predictions as long as we were informally well enough acquainted with the second of each of these pairs of situations. Furthermore, in both cases there would be some nontrivial empirical claim involved. For instance, in these cases it would be the claim that perception is generally veridical or that what we see generally corresponds to what we know to be the case. Although these are real empirical claims, no one takes them to have the theoretical significance that is attributed to corresponding theories of imagery, even though both may in fact have the same underlying basis.

Of course some models of imagery appear to go beyond such mere metaphors. For example, Kosslyn and Shwartz (1977) actually have a computer model of imagery that accounts for a very wide range of experimental findings. However, as we suggested in referring to its use of the blink transformation, unless the model incorporated a greater number of principled constraints, it is much too easy for it to accommodate any finding. It can do this because the model is, in fact, just a simulation of some largely commonsense ideas about what happens when we image. That is why the principles it appeals to are invariably stated in terms of properties of the represented domain (e.g., a principle such as *images must be transformed through small angles* clearly refers to what is being represented, since images themselves do not actually have orientations, as opposed to representing them). Yet this is how we often explain things in informal, everyday terms. We say that we imagine things in a certain way, because that is the way they really are. As I have already remarked, a theory of the underlying process should account for how imagery can come to have this character, not use this very property as an explanatory principle.

Another consequence of the model's being a simulation of commonsense views is that anything that could be stated informally as a description of what happens in the mind's eye can easily and naturally be accommodated in the model. For example, in arguing against the expectation or demand explanation of their scanning results, Kosslyn et al. (1979) cite one subject who said that he or she thought objects close together would take longer to image because it would be harder to see them or tell them apart. This is exactly the sort of process that could very easily be accommodated by the model. All that has to be done is to make the grain of the surface display whatever size is required to produce this effect. There is nothing to prevent this sort of tuning of the model to fit each situation. Such properties therefore have the status of free parameters, and unfortunately there is no limit to how many such parameters may be implicit in the model. Similarly, in our experiment (in which subjects judged the compass bearing of one place relative to a second) we found a negative correlation for some of the subjects between reaction time and distance between focal points on the imagined map. The computer model would have little difficulty accommodating that result if it turned out to be a general finding. In fact it is hard to think of any result which could not be naturally accommodated—including, as

Kosslyn et al. (1980) themselves suggest, the possibility of representing non-Euclidean space. What is crucial is not merely that such results could be accommodated, but that this could be done without violating any fundamental design criteria of the model and without threatening any basic principle of its operation—without, for example, violating any constraints imposed by the hypothesized surface display. What this amounts to is that the really crucial aspects of the model have the status of free parameters rather than structural constants.

Now to some extent Kosslyn appears to recognize this flexibility in his model. In the last section of Kosslyn et al. (1979), the authors insist that the model ought to be evaluated on the basis of its heuristic value. I agree with that proposal. The ad hoc quality characteristic of early stages of some scientific modeling may well be unavoidable. However we should make every effort to be realistic and rid ourselves of all the attending illusions. One of the illusions that goes with this way of thinking about imagery is that there is an essential core of the model that is not merely heuristic but is highly principled and highly constrained. That core is contained in the postulated properties of the surface display (or in the "cathode ray tube proto-model"). This immutable core involves the assumption that there is an internal display medium with intrinsic geometrical (or geometry-analogue) properties. For example, Pinker (1980) claims that the "array structure" captures a set of generalizations about images. But without knowing which properties of the array structure are doing the work, and whether such properties can be altered by changes in what the subject believes, this sort of capturing of generalizations may be no better than merely listing them. We need to know what it is about the intrinsic character of arrays that requires them to have the properties that Pinker suggests (e.g., "that they represent shape, size, and locations implicitly in an integral fashion, that they are bounded in size and grain, that they preserve interpoint distances" p. 148). If there is nothing apart from stipulation that requires arrays to have these properties, then each of these properties is precisely a free parameter.

For example, if the facts supported the conclusion that size, shape, and orientation were naturally factored apart in one's mental representation (as I believe they do), or that grain size is nonhomogeneous and varies as a function of what the subject believes the referent situation to be like (e.g., how brightly lit, how detailed in its design, how important different features are to the task at hand), does anybody believe for one moment that this would undermine the claim that an array structure was being used? Clearly all that it would require is some minor adjustment in the system (e.g., making resolution depend on additional features of the image, allowing the cognitive process to access an orientation parameter in memory and so on). But if that is the case, then it is apparent that claiming an array structure places no constraints on the sorts of phenomena that can be accommodated—that is, that properties of this structure are *free empirical parameters*. Thus, although one may have the impression that there is a highly constraining core assumption that is crucial to the predictive success of the model, the way this impression is maintained is simply by giving the rest of the system enough degrees of freedom to overcome any effort to empirically reject that core assumption. So, whereas the intuitive appeal of the system continues to hang on the unsupported view that properties of imagery are determined by the intrinsic properties of an internal display medium, its predictive power may in fact come entirely from a single empirical hypothesis of imagery theory (viz., the perception metaphor).

The phenomenon of having the real appeal of a theoretical system come from a simplified (and strictly false) view of the system while its predictions come from more complex (and more ad hoc) aspects is commonplace in science. In fact, even the initial success of the Copernican world view might be attributable to such a characteristic. Copernicus published his epoch-making proposal, *de Revolutionibus*, in two volumes. The first volume showed how the solar-centered system could in principle elegantly handle certain aspects of stellar and planetary motions (involving reverse movements) without the necessity of such ad hoc devices as epicycles.

In the second volume Copernicus worked out the details of his system for the more comprehensive case. That required reintroducing the ad hoc mechanisms of epicycles. In fact, Copernicus's system only did away with the five major Ptolomaic epicycles and retained all the complexity associated with the larger number of minor ones needed to make the theory fit the observations. Of course in time his system was vindicated because the discovery of the general principle of gavitation made it possible to subsume the otherwise ad hoc mechanisms under a universal law and hence to remove that degree of freedom from the theory.

The lesson for imagery is clear enough. In order for a theory of imagery to be principled it is necessary to locate the knowledge-independent functional properties correctly. We must be critical in laying a foundation of cognitively impenetrable functions to serve as the basic architecture of a formal model. The reason is not simply that in this way we can get to the most primitive level of explanation: It is rather that we can only get a principled and constrained (and therefore not ad hoc) model if we first fix those properties which are the basic functional capacities of the system. This does not mean, of course, that we must look to biology to provide us with a solution (though we can use help from all quarters), because the fixed functional capacities can be inferred behaviorally and specified functionally, as they are when the architecture of computers is specified. But it does mean that unless we set ourselves the goal of establishing the correct functional architecture or medium in order to properly constrain our models in the first place, we could well find ourselves in the position of having as many free parameters as we have independent observations.

Reference Notes

1. Bannon, L. *An investigation of image scanning.* Unpublished doctoral dissertation, Department of Psychology, University of Western Ontario, London, Canada, 1981.
2. Spoehr, K. T., & Williams, B. E. *Retrieving distance and location information from mental maps.* Paper presented at the 19th annual meeting of the Psychonomic Society, San Antonio, Texas, November 1978.
3. Kosslyn, S. M., Reiser, B. J., Farah, M., & Fliegel, S. L. *Generating visual images.* Unpublished manuscript (submitted for publication).

References

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

Arbib, M. *The metaphorical brain.* New York: Wiley, 1972.

Attneave, F. How do you know? *American Psychologist*, 1974, *29*, 493–499.

Chomsky, N. Rules and representations. *The Behavioral and Brain Sciences*, 1980, *3*(1), 1–62.

Finke, R. A. The functional equivalence of mental images and errors of movement. *Cognitive Psychology*, 1979, *11*, 235–264.

Fodor, J. A. The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy*, 1968, *65*, 627–640.

Fodor, J. A. *The language of thought.* New York: Thomas Y. Crowell, 1975.

Fodor, J. A. Methodological solipsism considered as a research strategy for cognitive psychology. *The Behavioral and Brain Sciences*, 1980, *3*(1), 63–110.

Forster, K. I. Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms.* Amsterdam: North-Holland, 1976.

Fraisse, P. *The psychology of time.* New York: Harper & Row, 1963.

Ghiselin, B. *The creative process.* New York: New American Library, 1952.

Hays, J. R. On the function of visual imagery in elementary mathematics. In W. G. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973.

Harman, G. *Thought.* Princeton, N.J.: Princeton University Press, 1973.

Haugeland, J. The nature and plausibility of cognitivism. *The Behavioral and Brain Sciences*, 1978, *2*, 215–260.

Hebb, D. O. Concerning imagery. *Psychological Review*, 1968, *75*, 466–477.

Hinton, C. H. *The fourth dimension.* London: George Allen & Unwin, 1906.

Howard, I. P. Recognition and knowledge of the water-level principle. *Perception*, 1978, *7*, 151–160.

Kosslyn, S. M. Scanning visual images: Some structural implications. *Perception & Psychophysics*, 1973, *14*, 90–94.

Kosslyn, S. M. The information represented in visual images. *Cognitive Psychology*, 1975, *7*, 341–370.

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*, 46–60.

Kosslyn, S. M., & Shwartz, S. P. A data driven simulation of visual imagery. *Cognitive Science*, 1977, *1*, 265–296.

Kosslyn, S. M., Pinker, S., Smith, G., & Shwartz, S. P. On the demystification of mental imagery. *The Behavioral and Brain Sciences*, 1979, *2*(3), 535–581.

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

Mitchell, D. B., & Richman, C. L. Confirmed reservations: Mental travel. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 58–66.

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

Newell, A., & Simon, H. A. Computer science as empirical inquiry. *Communications of the Association for Computing Machinery*, 1976, *19*, 113–126.

Paivio, A. U. Images, propositions and knowledge. In J. M. Nicholas (Ed.), *Images, perception, and knowledge.* Dordrech Holland: Reidel, 1977.

Palmer, S. F. Fundamental aspects of cognitive representation. In E. H. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, N.J.: Erlbaum, 1978.

Pinker, S. Explanations in theories of language and of imagery. *The Behavioral and Brain Sciences*, 1980, *3*(1), 147–148.

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 and artificial intelligence. In W. Savage (Ed.), *Perception and cognition: Issues in the foundations of psychology.* Minneapolis: University of Minnesota Press, 1978.

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

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

Pylyshyn, Z. W. Cognitive representation and the process-architecture distinction. *The Behavioral and Brain Sciences*, 1980, *3*(1), 154–169. (a)

Pylyshyn, Z. W. Computation and cognition: Issues in the foundations of cognitive science. *The Behavioral and Brain Sciences*, 1980, *3*(1), 111–132. (b)

Pylyshyn, Z. W., Elcock, E. W., Marmor, M., Sander, P. Explorations in perceptual-motor spaces. In *Proceedings of the second international conference of the Canadian Society for Computational Studies of Intelligence.* Toronto, Canada: University of Toronto, Department of Computer Science, 1978.

Richman, C. L., Mitchell, D. B., & Reznick, J. S. Mental travel: Some reservations. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 13–18.

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

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

Shulman, G. L., Remington, R. W., & McLean, J. P. Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *15*, 522–526.

Received October 11, 1979
Revision received January 20, 1980 ■