Mental Imagery: In search of a theory*

Zenon W. Pylyshyn
Rutgers Center for Cognitive Science
Rutgers University, Busch Campus
Piscataway, NJ 08854

ABSTRACT

It is widely accepted that there is something special about reasoning by using mental images. The question of how it is special, however, has never been satisfactorily spelled out, despite over thirty years of research in the post-behaviorist tradition. This article considers some of the general motivation for the assumption that entertaining mental images involves inspecting a picture-like object. It sets out a distinction between phenomena attributable to the nature of mind, to what is called the cognitive architecture, and ones that are attributable to tacit knowledge used to simulate what would happen in a visual situation. With this distinction in mind the paper then considers in detail the widely held assumption that in some important sense images are spatially displayed or are depictive, and that examining images uses the same mechanisms that are deployed in visual perception. I argue that the assumption of the spatial or depictive nature of images is only explanatory if taken literally, as a claim about how images are physically instantiated in the brain, and that the literal view fails for a number of empirical reasons – e.g., because of the cognitive penetrability of the phenomena cited in its favor. Similarly, while it is arguably the case that imagery and vision involve some of the same mechanisms, this tells us very little about the nature of mental imagery and does not support claims about the pictorial nature of mental images. Finally I consider whether recent neuroscience evidence clarifies the debate over the nature of mental images. I claim that when such questions as whether images are depictive or spatial are formulated more clearly, the evidence does not provide support for the picture-theory over a symbol structure theory of mental imagery. Even if all the empirical claims were true, they do not warrant the conclusion that many people have drawn from them; that mental images are depictive or are displayed in some (possibly cortical) space. Such a conclusion is incompatible with what is known about how images function in thought. We are then left with the provisional counterintuitive conclusion that the available evidence does not support rejection of what I call the “null hypothesis”; namely, that reasoning with mental images involves the same form of representation and the same processes as that of reasoning in general, except that the content or subject matter of thoughts experienced as images includes information about how things would look.
1 Why is there a problem about mental imagery?

1.1 The pull of subjective experience

Cognitive science is rife with ideas that offend our intuitions. It is arguable that nowhere is the pull of the subjective stronger than in the study of perception and mental imagery. It is not easy for us to take seriously the proposal that the visual system creates something like symbol structures in our brain since it seems intuitively obvious that what we have in our mind when we look out onto the world, as well as when we close our eyes and imagine a scene, is something that looks like the scene, and hence whatever it is that we have in our heads must be much more like a picture than a description. Though we may now that this cannot be literally the case, that it would do no good to have an inner copy of the world, this reasoning appears to be powerless to dissuade us from our intuitions. Indeed, the way we describe how it feels to imagine something shows the extent of the illusion; we say that we seem to be looking at something with our “mind’s eye”. This familiar way of speaking reifies an observer, an act of visual perception, and a thing being perceived. All three parts of this equation have now taken their place in one of the most developed theories of mental imagery (Kosslyn 1994), which refers to a
“mind’s eye” and a “visual system” that examines a “mental image” located in a “visual buffer”. Dan Dennett has referred to this view picturesquely as the “Cartesian Theater” view of the mind (Dennett 1991) and I will refer to it as the “picture theory” of mental imagery.

There has been a tradition of analyzing this illusion in the case of visual perception, going back to Descartes and Berkeley (it also appears in the 17th century debate between Arnaud and Malebranche – see Slezak submitted), and revived in modern times by (Gibson 1966), as well as computationalists like (Marr 1982). More recently (O’Regan 1992; O’Regan & Noë 2002) have argued against the intuitive picture-theory of vision on both empirical and theoretical grounds. Despite the widespread questioning of the intuitive picture view in visual perception, this view remains very nearly universal in the study of mental imagery (with such notable exceptions as Dennett 1991; Rey 1981; Slezak 1995); (see also the critical remarks by Fodor 1975; Hinton 1979; Pylyshyn forthcoming; Thomas 1999).

Why should this be so? Why do we find it so difficult to accept that when we “examine our mental image” we are not in fact examining an inner state, but rather are contemplating what the inner state is about – that is, some possible state of the visible world – and therefore that this experience tells us nothing about the nature and form of the representation? Philosophers have referred to this displacement of the object of thought from the (possible) world to a mental state as the “intentional fallacy” and it has much of cognitive science in its grip still.

What I try to do in this paper is show that we are not only deeply deceived by our subjective experience of mental imagery, but that the evidence we have accumulated to support what I call the “picture theory” of mental imagery is equally compatible with a much more parsimonious view, namely that most of the phenomena in question (but not all – see below) are due to the fact that the task of “imaging” invites people to simulate what they believe would happen if they were looking at the actual situation being visualized. I will argue that the alternative picture theory, or depiction-theory, trades so heavily on a systematic ambiguity between the assumption of a literal picture and the much weaker assumption that visual properties are somehow encoded. I will also argue that recent evidence from neuroscience (particularly the evidence of neural imaging) brings us no closer to a plausible picture theory than we were before this evidence was available.

1.2 The imagery debate: What was it about?

There has been a great deal of discussion in the past 30 years about “the imagery debate.” Many people even believe that the debate has, at least in general outline, been put to rest because we now have hard evidence from neuroscience showing what (and where) images are (see, e.g., Kosslyn 1994; and the brief review in Pylyshyn 1994a). But if one looks closer at the “debate” one finds that what people think the debate is about is very far from univocal. For example, some people think that the argument that has been settled is whether images, whatever their nature, are fundamentally different from the form of representation involved in other kinds of reasoning, whether there are two different systems of mental codes. For others it is the question of whether images have certain particular properties – e.g. whether they are spatial, or depictive, or analogue. Others feel that the question that has been settled is whether imagery “involves” the visual system. I will argue that none of these claims has been sufficiently well posed to admit of a solution. In this paper I will concentrate primarily on a particular class of theory of mental imagery, which I refer to as “picture theories” and will consider other aspects of the “debate” only insofar as they bear on the alleged pictorial nature of images.

In this article I defend the provisional view, which I refer to as the “null hypothesis,” that at the relevant level of analysis – the level appropriate for explaining the results of many experiments on mental imagery – the process of imagistic reasoning involves the same mechanisms and the same forms of representation as are involved in general reasoning, though with different content or subject matter. This hypothesis claims that what is special about image-based thinking is that it is typically concerned with a certain sort of content or subject matter, such as optical, geometrical, or what we might call the appearance-properties of the things we are thinking about. If so, nothing is gained by attributing a
special format or special mechanisms to mental imagery. While the validity of this null hypothesis remains an open empirical question, what is not open, I claim, is whether certain currently popular views can be sustained.

In the interest of full disclosure I should add that I don’t really, in my heart of hearts, believe that representations and processes underlying imagery are no different from other forms of reasoning. Nonetheless, I do think that nobody has yet articulated the specific way that images are different and that all candidates proposed to date are seriously flawed in a variety of ways that are interesting and revealing. Thus using the null hypothesis as a point of departure may allow us to focus more properly on the real differences between imagistic and other forms of reasoning.

1.3 Plan of the paper

Section 2 reviews some observations that have led many people to hold a picture theory of mental images (although a detailed discussion of what such a theory assumes is postponed until section 5). Section 3 introduces a distinction that is central to our analysis. It distinguishes two reasons why imagery might manifest the properties that are observed in experiments. One reason is that these properties are intrinsic to the architecture of the mental imagery system – they arise because of the particular brain mechanisms deployed in imagery. The other reason is that the properties are extrinsic to the mechanisms employed – they arise because of what people tacitly believe about the situation being imagined, which they then use to simulate certain behaviors that would occur if they were to witness the corresponding situation in reality. This distinction is then applied to some typical experiments on mental imagery where I argue that such experiments tell us little about special dedicated imagery mechanisms. Since section 4 discusses some material that has been published elsewhere, readers who have followed the “imagery debate” may wish to skim this section.

Section 5 discusses two widely held views about the nature of mental images (Kosslyn 1994); that images are “depictive” and that they are laid out in a “functional space”. I claim that the preponderance of evidence argues against the inherent spatial nature of mental images. An exception is evidence from experiments in which subjects project their images onto a visual scene. In this case I claim (section 5.3) that the use of visual indexes and focal attention provides a satisfactory explanation for how spatial properties are inherited from the observed scene, without any need to posit spatial properties of images. In section 0 I argue that the notion of a functional space is devoid of any explanatory power, since such a “space” is unconstrained and can have whatever properties one wishes to attribute to it (unless it is taken to be a simulation of a real spatial display as in the model described in Kosslyn, Pinker, Smith, & Shwartz 1979, in which case the underlying theory really is the literal picture theory). Section 6 discusses a claim that is assumed to be entailed by the depictive nature of images; namely, that information in an image is accessed through vision. Although there is evidence for some overlap between the mechanisms of imagery and those of vision, a close examination of this evidence shows that it does not support the assumption of a spatial display in either vision or imagery. Section 7 considers evidence from neuroscience, that many writers believe provides the strongest case for a picture theory. Here I argue that, notwithstanding the intrinsic interest of these findings, they do not support the existence of any sort of depictive display in mental imagery. Finally, section 8 closes with a brief discussion of where the “imagery debate” now stands and on the role of imagery in creative thinking.

2 What is special about image-based reasoning?

Imagery seems to follow principles that are different from those of intellectual reasoning and certainly beyond any principles to which we have conscious intellectual access. Imagine a baseball being hit into the air and notice the trajectory it follows. Although few of us could calculate the shape of this trajectory none of us has any difficulty imagining the roughly-parabolic shape traced out by the ball in this thought experiment. Indeed, we can often predict with considerable accuracy where the ball will land (certainly a properly situated professional fielder can). It is very often the case that by visualizing a
certain situation, we can predict the dynamics of physical processes that are beyond our ability to solve analytically. Is this because our imagery architecture inherently and automatically obeys the relevant laws of nature?

Opposing the intuition that one’s image unfolds according to some internal principle of natural harmony with the real world, is the obvious fact that it is you alone who controls your image. Perhaps, as (Humphrey 1951) once put it, viewing the image as being responsible for what happens in your imagining puts the cart before the horse. In the baseball example above, isn’t it equally plausible that the reason the imagined ball takes a particular path is that, under the right circumstances, you can recall having seen a ball inscribe such a path? Surely your image unfolds as it does because you, the image creator, made it do so. You can imagine things being pretty much any size, color or shape that you choose and you can imagine them moving any way you like. You can, if you wish, imagine a baseball sailing off into the sky or following some bizarre path, including getting from one place to another without going through intervening points, as easily as you can imagine it following a more typical trajectory. You can imagine all sorts of physically impossible things happening — and cartoon animators frequently do, to our amusement.

Some imagery theorists might be willing to concede that in imagining physical processes we must use our tacit knowledge of how things work, yet insist that the optical and geometrical properties of images are true intrinsic properties, despite that fact that the dynamic properties of images that are often cited in studies of mental images – properties such as mental rotation, mental scanning, or “representational momentum” discussed in sections 3.1 and 4. Nonetheless, the suggestion that the intrinsic properties of images are geometrical rather than dynamic makes sense both because spatial intuitions are among the most entrenched, and because there is evidence (Pylyshyn 1999) that geometrical and optical-geometrical constraints are built into the early-vision system, as so-called “natural constraints”. While we can easily imagine the laws of physics being violated, it seems nearly impossible to imagine the axioms of geometry and geometrical optics being violated. Try imagining a four-dimensional block or how a cube looks when seen from all sides at once or what it would look like to travel through a non-Euclidian space. However, before concluding that these examples illustrate the intrinsic geometry of images, consider whether your inability to imagine these things might not be due to your not knowing, in a purely factual way, how these things might look (that is, where edges, shadows and other contours would fall)? The answer is by no means obvious. It has even been suggested (Goldenberg & Artner 1991) that certain deficits in imagery ability resulting from brain damage, are a consequence of a deficiency in the patient’s knowledge about the appearance of objects. At the minimum we are not entitled to conclude from such examples that images have the sort of inherent geometrical properties that we associate with pictures.

We also need to keep in mind that notwithstanding one’s intuitions, there is reason to be skeptical about what one’s subjective experience reveals about the form of a mental image. After all, when we look at an actual scene we have the unmistakable subjective impression that our perceptual representation is of a detailed three-dimensional panoramic view, yet it has now been convincingly demonstrated that the information available to cognition from a single glance is extremely impoverished, sketchy and unstable and that very little is carried over across saccades (see, for example, Blackmore, Brelstaff, Nelson, & Troscianko 1995; Carlson-Radvansky 1999; Carlson-Radvansky & Irwin 1995; Intraub 1981; Irwin 1993; O'Regan 1992; O'Regan & Noë 2002; Rensink 2000a, 2000b; Rensink, O'Regan & Clark 1997; Rensink, O'Regan & Clark 2000; Simons 1996). Indeed, there is now considerable evidence that we visually encode very little in a visual scene unless we explicitly attend to the items in question and that we do that only if our attention or our gaze is attracted to it (Henderson & Hollingworth 1999), (although see O'Regan, Deubel, Clark, & Rensink 2000). There are remarkable demonstrations that when presented with alternating images, people find it extremely difficult to detect a difference between the two – even a salient difference in a central part of the image. This so-called change blindness phenomenon (Simons & Levin 1997) suggests that, notwithstanding our
phenomenology, we are nowhere near having a detailed internal display since the vast majority of information in a visual scene goes unnoticed and unrecorded. It would thus be reasonable to expect that our subjective experience of mental imagery would be an equally poor guide to the form and content of the information in our mental images.

3 Why images exhibit certain properties: Cognitive architecture or tacit knowledge?

Nobody denies that the content and behavior of our mental images can be the result of what we intend our images to show, what we know about how things in the world look and work, and the way our mind or our imagery system constrains us. The important question about mental imagery is; which properties and mechanisms are intrinsic to, or constitutive of having and using mental images, and which arise because of what we believe, intend, or attribute to the situation we are imagining.

The distinction between effects attributable to the intrinsic nature of mental mechanisms and those attributable to more transitory states, such as people’s beliefs, utilities, habits, or interpretation of the task at hand, is central not only for understanding the nature of mental imagery, but for understanding mental processes in general. Explaining the former kind of phenomena requires that we appeal to what has been called the cognitive architecture (Fodor & Pylyshyn 1988; Newell 1990; Pylyshyn 1980, 1984; Pylyshyn 1991a, 1996) – one of the most important ideas in cognitive science. It refers to the set of properties of mind that are fixed with respect to certain kinds of influences. In particular, the cognitive architecture is, by definition, not directly altered by changes in knowledge, goals, utilities or any other representations (e.g., fears, hopes, fantasies, etc). In other words when you find out new things or when you draw inferences from what you know or when you decide something, your cognitive architecture does not change. Of course, if as a result of your state of beliefs and desires you decide to take drugs or to change your diet or even to repeat some act over and over, this can result in changes to your cognitive architecture, but such changes are not a direct result of the changes in your cognitive state. A detailed technical exposition of the distinction between effects attributable to knowledge or other cognitive states and those attributable to the nature of cognitive architecture are beyond the scope of this article (although this distinction is the subject of extensive discussion in Pylyshyn 1984, Chapter 7). This informal characterization and the following example will have to do for present purposes.

Consider the following illustrative example. We have a box of unknown construction, and we discover that it exhibits particular systematic behaviors (discussed in, Pylyshyn 1984). The box emits long and short pulses according to the following pattern: pairs of short pulses most often precede single short pulses, except when a pair of long-short pulses occurs first. What is special about this example is that it illustrates a case where the observed behavior, though completely regular when the box is in its “ecological niche,” is not due to the nature of the box (to how it is constructed) but to an entirely extrinsic reason. The reason this particular pattern of behavior occurs can only be understood if we know that the pulses are codes, and the pattern is due to a regularity in what they represent, in particular, that the pulses represent English words spelled out in International Morse Code. The observed pattern does not reflect how the box is wired or its functional architecture; it is due entirely to a regularity in the way English words are spelled (the principle being that generally \(i\) comes before \(e\) except after \(c\)). Similarly, I have argued that in most of the core experiments on mental imagery – such as the mental scanning case described in section 4.1 – the pattern does not reveal the nature of the mental architecture involved in imagery, but reflects a principle that observers know governs the world being imagined. The reason that in under certain conditions the behavior of both the code box and the cognitive system does not reveal properties of its intrinsic nature (of its architecture) is that both are capable of quite different regularities if the world they were representing behaved differently. They would not have to change their architecture in order to change their behavior. The latter observation, concerning the plasticity of non-architectural properties of thought, is the key to a methodology I have called “cognitive penetrability” for deciding whether tacit knowledge or cognitive architecture is
responsible for some particular observed regularity (see section 3.2).

In interpreting the results of imagery experiments, it is clearly important to distinguish between cognitive architecture and tacit knowledge as possible causes. Take the following example. You are asked what color you see if you look through a yellow filter superimposed on a blue filter. The way that many of us would go about solving this problem, if we did not know the answer as a memorized fact, is to imagine a yellow filter and a blue filter being superimposed; we generally use the “imagine” strategy when we want to solve a problem about how certain things look. What color do you see in your image when the two filters are overlapped? Now ask yourself why you see that color in your mind’s eye rather than some other color? Some people (e.g., Kosslyn 1981) have argued that the color you see follows from a property of imagery, presumably some property of how colors are encoded and displayed in images. But since there can be no doubt that you can make the overlapping part of the filters be any color you wish, it can’t be that the image format or the architecture involved in representing colors is responsible. What else can it be? It seems clear in this case that the color you “see” depends on your tacit knowledge of the principles of color mixing or a recollection of how these particular colors combine (having seen something like them in the past). In fact people who do not know about subtractive color mixing generally give the wrong answer: mixing yellow light with blue light produces white light, but overlapping yellow and blue filters allows green light through.

When asked to do this exercise (as reported in Kosslyn 1981), some people claim that they “see” a color that is different from the one they report when they are simply asked to say (without using imagery) what would happen. Results such as this have made people leery of accepting the tacit knowledge explanation. There are indeed many cases where people report a different result when using mental imagery than when asked to merely answer the question without using their image. It is not clear what moral ought to be drawn from this, however, since it is a general property of reasoning that the way the question is put and the reasoning strategy that is used to get to the answer can affect the outcome. Knowledge can be organized and accessed in many different ways (see section 4.3 for more on the relevance of this to mental imagery studies). Indeed, it need not be accessed at all if it seems like more work that it is worth. For example, consider the following analog of the color-mixing task. Close your eyes and imagine someone writing the following on a blackboard: “759 + 356 = ”. Now, imagine that the person continues writing on the board. What number can you “see” being written next? People may give different answers depending on whether they believe that they are supposed to work it out or whether in the interest of speed they are supposed to guess or merely say whatever comes to mind. Each of these is a different task. Even without a theory of what is special about visual imagery, we know that the task of saying what something would look like can be a different task from the task of solving a certain intellectual puzzle about colors or numbers.

In most of the cases studied in imagery research, it would be odd if the results did not come out the way picture theorists predict. For if the results were inconsistent with the picture-theory, the obvious explanation would be that subjects either did not know how things would work in reality or else they misunderstood the instructions to “imagine x”. For example if you were asked to imagine in vivid detail, a performance of the Minute Waltz, the failure of the imagined event to take approximately one minute would simply indicate that you had not carried out the task you were supposed to. Since taking roughly one minute is constitutive of a real performance, it is natural to assume it to be indicative of a realistic imaginary re-creation of such a performance.

3.1 What knowledge is relevant to the tacit knowledge explanation?

The concept of tacit knowledge plays an important role in cognitive science (see, for example, Fodor 1968), though it has frequently been maligned because it has to be inferred indirectly. Such knowledge is called “tacit” because it is not always explicitly available for, say, answering questions. There may nonetheless be independent evidence that such knowledge exists. This is a point that has been made forcibly in connection with tacit knowledge of grammar or of social conventions, which typically also
cannot be articulated by members of a linguistic or social group, even though violations are easily detected. In our case the role of tacit knowledge can sometimes be detected using the criterion of cognitive penetrability, discussed below.

Not only is the notion of tacit knowledge often misunderstood, but in the case of explaining mental imagery results, the kind of tacit knowledge that is relevant has also been widely misunderstood. The only knowledge that is relevant to the tacit knowledge explanation is knowledge of what things would look like to subjects in situations like the ones in which they are to imagine themselves. Many writers have mistakenly assumed that the tacit knowledge explanation refers to one of several other kinds of knowledge. For example, although tacit knowledge of what results the experimenter expects (sometimes referred to as “experimenter demand effects”) is always an important consideration in psychological experiments (and may be of special concern in mental imagery experiments; see Banks 1981; Intons-Peterson 1983; Intons-Peterson & White 1981; Mitchell & Richman 1980; Reed, Hock & Lockhead 1983; Richman, Mitchell & Reznick 1979) it is not the knowledge that is relevant to the tacit knowledge explanation, as some have assumed (Finke & Kurtzman 1981b). Nor is it knowledge of such things as how the visual system works. It is not relevant to the tacit knowledge explanation that people are unlikely to know how their visual system or the visual brain works (as Farah 1988, has assumed). It is also not the knowledge people might have of what results to expect from experiments on mental imagery (as assumed by Denis & Carfantan 1985). Denis & Carfantan studied “people’s knowledge about images” and found that people often failed to correctly predict what would happen in experiments such as mental scanning. But these sorts of questions invite respondents to consider their folk psychological theories to make predictions about psychological experiments. They do not reflect tacit knowledge of what it would look like if the observers were to see a certain event happening in real life. The tacit knowledge claim is simply the claim that when subjects are asked to “imagine x” they use their knowledge of what “seeing x” would be like (as well as their other psychophysical skills, such as estimating time-to-collision) and they simulate as many of these effects as they can. Whether a subject has this sort of tacit knowledge cannot always be determined by asking them, and certainly not by testing them for their knowledge of psychology!

Notwithstanding the importance of tacit knowledge explanations of imagery phenomena, it remains true that not all imagery results are subject to this criticism. Even when tacit knowledge is involved, there is often than one reason for the observed phenomena. An example in which tacit knowledge may not be the only explanation of an imagery finding can be found in (Finke & Pinker 1982). The example concerns a particular instance of mental scanning (one in which it takes more time to judge that an arrow points to a dot when the dot is further away). Finke and Pinker argued that these results could not have been due to tacit knowledge because, even though subjects correctly predicted that judgments would take more time when the dots were further away, they failed to predict that the time would actually be longer for the shortest distance used in the study. But this was a result that even the authors failed to anticipate, because the aberrant short-distance time was most likely due to some mechanism (perhaps attentional crowding) different from the one that caused the monotonic increase of time with distance.

Another example in which tacit knowledge does not account for some aspect of an imagery phenomenon is in what has been called “representational momentum”. It was shown that when subjects observe a moving object and are asked to recall its final position from memory, they tend to misremember it as being displaced forward. (Freyd & Finke 1984) attributed this effect to a property of the imagery architecture. On the other hand, (Ranney 1989) suggested that the phenomenon may actually be due to tacit knowledge. It seems that at least some aspects of the phenomenon may not be attributable to tacit knowledge (Finke & Freyd 1989). But here again there are other explanations besides tacit knowledge or image architecture. In this particular case there is good reason to think that part of the phenomenon is actually visual. There is evidence that the perception of the location of moving objects is ahead of the actual location of such objects (Nijhawan 1994). Eye movement studies have also shown that gaze precedes the current location of moving objects in an anticipatory fashion (Kowler 1989, 1990). Thus even though the general phenomenon, involving imagined motion, may be
attributable to tacit knowledge, the fact that the moving stimuli are presented visually may result in the phenomena also being modulated by the visual system. The general point in both these examples is that even in cases where tacit knowledge is not the sole determiner of a result in an imagery experiment, the phenomena in question need not reveal properties of the architecture of the imagery system. They may be due to properties of the visual system, the memory system, or a variety of other systems that might be involved.

3.2 Methodological note: cognitive penetrability as a litmus

How is it possible to tell whether certain imagery effects reflect the nature of the imagery architecture or the person’s tacit knowledge? In general, methodologies for answering questions about theoretical constructs are limited only by the imagination of the experimenter. Typically they involve convergent sources of evidence and independent theoretical motivation. One theoretically motivated diagnostic, discussed at length in (Pylyshyn 1984), is to test for the cognitive penetrability of the observations. This criterion is based on the assumption that if a particular pattern of observations arises because people are simulating a situation based on their tacit beliefs, then if we alter their beliefs or their assumptions about the task, say by varying the instructions, the pattern of observations may change accordingly, in ways that are rationally connected with the new beliefs. So, for example, if we instruct a person on the principles of color mixing we would expect the answer to the imaginary color-mixing question discussed above to change appropriately. We will see other examples of the use of this criterion throughout this article (especially the examples in section 4).

Not every imagery-related phenomenon that is genuinely cognitively impenetrable provides evidence for the nature of mental images or their mechanisms. Clearly many beliefs are resistant to change by merely being told that they are false. Nonetheless this criterion has proven useful in identifying parts of the visual system that constitute what is called early vision (Pylyshyn 1999). Cognitive penetrability remains a necessary but not sufficient condition for a pattern being due to the architecture of the imagery system.

4 Problem-solving by “mental simulation”: Some examples

The idea that what happens in certain kinds of problem solving can be viewed as off-line simulation has had a recent history in connection not only with mental imagery (Currie 1995), but also with other sorts of problems in cognitive science (Klein & Crandall 1995). But even if we grant that the “simulation mode” of reasoning is used in various sorts of problem solving, the question still remains; what does the real work in solving the problem by simulation – a special property of images (that is, the architecture of the image system) or tacit knowledge?

In what follows I will sketch a number of influential experimental results often cited in support of the picture theory, and compare explanations given in terms of inherent properties of the image with those given in terms of simulation based on tacit knowledge.

4.1 Scanning mental images

Probably the most cited result in the entire repertoire of research motivated by the picture-theory is the image-scanning phenomenon. Not only has this experimental paradigm been used dozens of times, but various arguments about the “metrical” or spatial nature of mental images, as well as arguments about such properties of the mind’s eye as its “visual angle,” rest on this phenomenon. Indeed, it has been referred to as a “window on the mind” (Denis & Kosslyn 1999).

The finding is that it takes longer to “see” a feature in a mental image that is further away from the place in the image an observer was initially focusing upon. So, for example, if you are asked to imagine a dog and inspect its nose and then to “see” what its tail looks like it will take you longer than if you were asked to first inspect its hind legs. Here is an actual experiment reported in (Kosslyn, Ball & Reiser 1978). Subjects were asked to memorize a map such as the one in Figure 1. They were then
asked to imagine the map and to focus their attention on one place — say the “church”. In a typical experiment (there are many variants of this basic study) the experimenter says the name of a second place (say, “beach” or “tree”) and subjects are asked to examine their image and to press a button as soon as they can “see” the second named place on their image of the map. What many researchers have found consistently is that the further away the second place is from where the subject is initially focused, the longer it takes to “see” the second place in the image.

From this scanning result most researchers have concluded that larger map distances are represented by greater distances in image space. In other words, the conclusion that is drawn from this kind of experiment is that mental images have spatial properties – that is, they have spatial magnitudes or distances, as opposed to just encoding such properties in some unspecified manner. This is a strong conclusion about cognitive architecture. It says, in effect, that the symbolic code idea that forms the foundation of computational theories does not apply to mental images. In a symbolic encoding two places can be represented as being further away just the way we do it in language; by saying the places are, say, \( n \) meters from one another. But the representation of larger distances is not itself in any sense larger.

Is this strong conclusion about the metrical property of mental images warranted? Does the difference in scanning time reveal a property of the architecture or a property of what is represented? Notice how this distinction exactly parallels the situation in the color-mixing example discussed earlier. There we asked whether a particular observation revealed a property of the architecture or a property of what people know or believe – a property of the represented situation of which they have tacit knowledge. To answer this question for the scanning experiment we need to determine whether the pattern of increasing reaction time arises from a fixed capacity of the image-encoding or image-examining system or whether it can be altered by changing subjects’ understanding of the task or the beliefs that they hold about what it would be like to examine a real map; whether it is cognitively penetrable.

This is a question to be settled in the usual way – by careful analyses and experiments. But even before we do the experiment there is reason to suspect that the time-course of scanning is not a property of the cognitive architecture. Do the following test on yourself. Imagine that there are lights at each of the places on your mental image of the map. Imagine that a light goes on at, say, the beach. Now imagine that this light goes off and one comes on at the lighthouse. Did you need to scan your attention across the image to see the light come on at the lighthouse? Liam Bannon and I repeated the scanning experiment (see the description in Pylyshyn 1981) by showing subjects a real map with lights at the

![Figure 1: Map to be learned and imaged in one’s “mind’s eye” to study mental scanning](image)
target locations, much as I just described. We allowed the subjects to turn lights on and off. Whenever a light was turned on at one location it was simultaneously extinguished at another location. Then we asked subjects to imagine that very map and to indicate (by pressing a button) when a light was on and they could “see” the illuminated place in their image. The time between button presses was recorded and its correlation to the distances between illuminated places on the map was computed. We found that there was no relation between distance on the imagined map and time. You might think: Of course there was no increase in time with increasing distance, because subjects were not asked to imagine scanning that distance. But that’s just the point: You can imagine scanning over the imagined map if you want to, or you can imagine just hopping from place to place on the imaginary map. If you imagine scanning, you can imagine scanning fast or slow, at a constant speed or at some variable speed, or scanning part way and then turning back or circling around! You can, in fact, do whatever you wish since it is your image.  

At least you can do these things to the extent that you can create the phenomenology or the experience of them and providing you are able to generate the relevant measurements, such as the time you estimate it would take to get from point to point.

Whether or not you choose to simulate a certain temporal pattern of events in the course of answering a question may also depend in part on whether simulating that particular pattern seems to be relevant to the task. It is not difficult to set up an experimental situation in which simulating the actual scanning from place to place does not appear so obviously relevant to solving a particular problem. For example, we ran the following experiment that involved extracting information from an image (Pylyshyn 1981). Subjects were asked to memorize a map and to refer to their image of the map in solving the problem. As in the original (Kosslyn et al. 1978) studies, subjects had to first focus on one place on their imagined map and then to “look” at a second named place. The experiment differed from the original study, however, in that the task was to indicate the compass direction from the second named place to the previously focused place. This direction-judgment task requires that the subject make a judgment from the perspective of the second place, so it requires focusing at the second place. Yet in this experiment, the question of how you get from the first place to the second place on the map was far less prominent than it was in the “tell me when you can see X” task. In this study we found that the distance between places had no effect on the time taken to make the response. Thus it seems that the effect of distance on reaction time is cognitively penetrable.

Not only do observers sometimes move their attention from one imagined object to another without scanning through the space between them, but we have reason to believe that they cannot move their attention continuously through empty imagined space (see section 0 for a brief description of the relevant study).

4.2 The “size” of mental images

Another study closely related to the mental scanning paradigm, is one where it is found that it takes more time to report some visual detail of an imagined object if the object is imagined to be small, than if it is imagined to be large (e.g., it takes longer to report that a mouse has whiskers if the mouse is imagined as tiny, than if it is imagined as huge). This seems like a good candidate for a tacit knowledge explanation, since when you actually see a small object you know that you can make out fewer of its details due to the limited resolution of your eye. So if you are asked to imagine something small, you are likely to imagine it as having fewer visible details than if you are asked to imagine it looming large directly in front of you, whatever form of representation that may involve.

The original picture-theory view of this result is problematic in any case. What does it mean for your image to be “larger”? Such a notion is meaningful only if the image has a real size or scale. If, as in our null hypothesis, the information in the image is in the form of a symbolic description, then size has no literal meaning. You can think of something as large or small, but that does not make some thing in your head large or small. On the other hand, which details are represented in your imagination does
have a literal meaning: You can put more or less detail into your active representation. Inasmuch as
the task of imagining the mouse as “small” entails that you imagine it having fewer visible details, the
result is predictable without any notion of real scale applying to the image.

The obvious test of this proposal is to apply the criterion of cognitive penetrability. Are there
instructions that can ameliorate the effect of the “image size” manipulation, making details easier to
report in small images than in large ones and vice versa? Could you imagine a small but extremely high
resolution and detailed view of an object, in contrast to a large but low-resolution or fuzzy view that
lacks details? I know of no one who has bothered to carry out an experiment that asks subjects to, say,
report details from a large blurry image and then from a small clear one. What if such an experiment
were done and showed that it is quicker to report details from a large blurry object than a small clear
one? The strangeness of such a possibility should alert us to the fact that what is going wrong lies in
what it means to have a blurred versus a clear image. Such results would be incompatible with what
happens in seeing. If it took longer to see fine details in a real large object there would have to be a
reason for it, such as that you were seeing it through a fog or out of focus. Thus so long as examining a
visual image means simulating what it is like to see something, the results must be as reported; how
could studies involving different sized mental images, or blurred versus clear images, fail to show that
they parallel the case of seeing, unless subjects not understanding the instructions (e.g., did not
understand the meaning of “blurry”)? The same goes for the imagery analogue of any property of
seeing of which observers have some tacit knowledge or recollection. Thus it applies to the findings
concerning the acuity profile of imagery, which approximates that of vision (Finke & Kosslyn 1980).
Observers do not need to have articulated scientific knowledge of visual acuity; all they need is to
remember roughly how far into the periphery of their visual field things can go before they become hard
to see, and it is not surprising that this is easier to do while turning your head (with eyes closed) and
pretending to be looking at objects in your periphery (which is how these studies were done).

4.3 Mental “paper folding”

There are many reasons why one might use a “simulation mode” strategy in answering a question;
reasons that have nothing to do with the spatial nature of imagery, and sometimes not even because of
what tacit knowledge is available. For example, to answer the question: What is the fourth (or n’th)
letter in the alphabet after “M,” people normally have to go through the alphabetical sequence (and it
takes them longer the larger the value of n). Similarly, the findings reported by (Shepard & Feng 1972)
are easily understood if one considers how the relevant knowledge is organized. In their experiment,
subjects are asked to mentally fold pieces of paper, such as shown in Figure 2, and to report whether the
arrows marked on the paper would touch one another. They found that the more folds it would require
to actually fold the paper and see whether the arrows coincide, the longer it takes to imagine doing so.
From this they concluded that working with images parallels working with real objects.

![Figure 2: Two of the figures used in the (Shepard & Feng 1972) experiment. The task is to imagine folding the paper (using the dark shaded square as the base) and say whether the arrows in these two figures coincide. The time it takes increases with the number of folds required.](image)
The question that needs to be asked about this task is the same as the question we asked in connection with the color mixing task: What is responsible for the relation between time taken to answer the question and the number of folds it would have taken in folding real paper? This time the answer is not simply that it depends on tacit knowledge, because in this case it is not just the content of the tacit knowledge that makes the difference. It is because of the knowledge that subjects have about paper folding that they can do the task at all. But in this case it appears that imagining making individual folds is required in order to get the answer, and one would presumably get the same result even if one were not trying to imagine folding the paper. It is hard to see how to answer to this question without imagining going through the sequence of folds. A plausible explanation for this, which does not appeal to special properties of a mental image system, is that the reason one has to imagine going through a sequence of individual folds is the same as the reason one had to go through a series of letters in the earlier alphabet example. The reason may have to do with how one’s knowledge of the effects of folding is organized. What we know about the effects of paper folding is just this: we know what happens when we make one fold. Consequently to determine what would happen in a task that requires 4 folds, we have to apply our one-fold-at-a-time knowledge four times. Recall the parallel case with letters: In order to determine the fourth letter after M is we have to apply the “next letter” rote knowledge four times. In both cases a person could, in principle, commit to memory such facts as what results from double folds of different types; or which letter of the alphabet occurs exactly \( n \) letters after a given letter. If that were how paper-folding knowledge was organized, the Shepard and Feng results might not hold. The important point is that once again the result tells us nothing about how the states of the problem are represented — or about any special properties of image representations. They tell us only what knowledge the person has and how it is organized.

The role played by the structure of knowledge is ubiquitous and may account for another common observation about the use of mental imagery in recall. We know that some things are easier to recall than others and that it is easier to recall some things when the recall is preceded by the recall of other things. Memory is linked in various intricate ways. In order to recall what you did on a certain day it helps to first recall what season that was, what day of the week it was, where you were at the time, and so on. (Sheingold & Tenney 1982; Squire & Slater 1975) and others have shown that one’s recall of distant events is far better than one generally believes because once the process of retrieval begins it provides clues for subsequent recollections. The reason for bringing up this fact about recall is that such sequential dependencies are often cited as evidence for the special nature of imagery (Bower 1976; Paivio 1971). Thus, for example, in order to determine how many windows are there in your home, you probably need to imagine each room in turn and look around to see where the windows are, counting them as you go. In order to recall whether someone you know has a beard (or glasses or red hair) you may have to first recall other aspects of what he or she looks like (that is, recall an image of them). Apart from the phenomenology of recalling an appearance, what is going on is absolutely general to every form of memory retrieval. Memory access is an ill understood process, but at least it is known that it has sequential dependencies and other sorts of access paths and that these paths are often dependant on spatial arrangements (which is why the “method of loci” works well as a mnemonic device).

4.4 Mental Rotation

One of the earliest and most cited results in the research on manipulating mental images is the “mental rotation” finding. (Shepard & Metzler 1971) showed subjects pairs of drawings of three-dimensional figures, such as those illustrated in Figure 3, and asked them to judge whether the two objects depicted in the drawings were identical, except for orientation. Half the cases were mirror reflections of one another (or the 3D equivalent, called enantiomorphs), and therefore could not be brought into correspondence by a rotation. Shepard and Metzler found that the time it took to make the judgment was a linear function of the angular displacement between the pair of objects depicted.
Figure 3. Examples similar to those used by (Shepard & Metzler 1971) to show “mental rotation.” The time it takes to decide whether two figures are identical except for rotation (a, b) or are mirror images (a, c) increases linearly as the angle between them increases.

This result has been universally interpreted as showing that mental images of the objects are “rotated” continuously and at constant speed in the mind and that this is, in fact, the means by which the comparison is made: We rotate one of the pair of figures until the two are sufficiently in alignment that it is possible to see whether they are the same or different. The phenomenology of the Shepard and Metzler task is clearly that we rotate the figure in making the comparison. I do not question either the phenomenology nor the description that what goes on in this task is “mental rotation.” But there is some question about what these results tell us about the nature of mental images. The important question is not whether we can or do imagine rotating a figure, but whether we solve the problem by means of the mental rotation. For mental rotation to be a mechanism by which the solution is arrived at, its utility would have to depend on some intrinsic property of images. As an example if it were the case that during mental rotation the figure moves as a rigid form through a continuum of angles, then mental rotation would be capitalizing in an intrinsic property of the image format.

Contrary to the general assumption, however, figural “rotation” is not a holistic process that operates on an entire figure, while the figure retains its rigid shape. Subjects in the original 3D rotation study (Shepard & Metzler 1971) examined both the target and the comparison figures together. In a subsequent study that monitored eye movements, (Just & Carpenter 1976) showed that observers look back and forth between the two figures, checking for distinct features. This point was also made using simpler 2D figures where it was found that observers concentrate on significant milestone features when carrying out the task (Hochberg & Gellman 1977), and that when such milestone features are available, no rotation effect is found. In studies reported in (Pylyshyn 1979) I showed that the apparent “rate of rotation” depends both on the complexity of the figure and on the complexity of the post-rotation comparison task (I used a task in which observers had to indicate whether a test figure, presented at various orientations, was embedded within the original figure). The dependence of the rotation speed on such organizational and task factors shows that whatever is going on in this case does not appear to consist in merely “rotating” one figure in a rigid manner into correspondence with the reference figure.

Even if the process of making the comparison in some sense involves the “rotation” of a represented shape, this tells us nothing about the form of the representation and does not support the view that the representation is pictorial. The proposal that a representation maintains its shape because of the inherent rigidity of the image while it is rotated cannot be literally true, notwithstanding the phenomenology. The representation is not literally being rotated; no codes or patterns of codes are being moved in a circular motion. At most what could be happening is that a representation of a figure is processed in such a way as to produce a representation of a figure at a slightly different orientation, and then this process is iterated (perhaps even continuously). There are probably good reasons, based on computational resource considerations, why the process might proceed by iterating parts of a form over successive small angles (thus causing the comparison time to increase with the angular disparity between the figures) rather than attempt the “rotation” in one step. For example, (Marr & Nishihara
1976) hypothesized what they called a primitive SPASAR mechanism, whose function was to compute the rotation of a simple dihedral vertex and determine its orthographic projections in a reference frame (a slightly different version that left out the details of the SPASAR mechanism, was later published in Marr & Nishihara 1978). This was an interesting idea that assumed a limited analogue operation that could be applied to one small feature of a representation at a time. Yet the Marr and Nishihara proposal did not postulate a pictorial representation, nor did it assume that a rigid configuration was maintained by an image in the course of its “rotation.” It hypothesized a simple primitive operation on parts of a structured representation in a response to a computational complexity issue.

Like the paper folding task discussed earlier, the mental rotation phenomenon is robust and not cognitively penetrable, and is not a candidate for a straightforward tacit knowledge explanation (as I tried to make clear in Pylyshyn 1979). Rather, the most likely explanation is one that appeals to the computational requirements of the task and general architectural (that is, working memory) constraints, and therefore applies regardless of the form of the representation. No conclusions concerning the format of image representations, or the form of their transformation, follow from the rotation results. Indeed these findings illustrate how treating the phenomenology as explanatory does not help us to understand why or how the behavior occurs.

5 Are images “depictive”?

5.1 Depiction and mandatory spatial properties of representations

It has frequently been suggested that images differ from structured descriptions in that the former stand in a special relationship to what they represent, a relationship referred to as depicting. One way of putting this is to say that in order to depict some state of affairs the representation needs to correspond to the spatial arrangement it represents the way that a picture does. One of the few people who have tried to be explicit about what this means is Stephen Kosslyn,[4] so I quote him at some length [Kosslyn, 1994 #880, p5):

“A depictive representation is a type of picture, which specifies the locations and values of configurations of points in a space. For example, a drawing of a ball on a box would be a depictive representation. The space in which the points appear need not be physical, such as on this page, but can be like an array in a computer, which specifies spatial relations purely functionally. That is, the physical locations in the computer of each point in an array are not themselves arranged in an array; it is only by virtue of how this information is “read” and processed that it comes to function as if it were arranged into an array (with some points being close, some far, some falling along a diagonal, and so on). In a depictive representation, each part of an object is represented by a pattern of points, and the spatial relation among these patterns in the functional space correspond to the spatial relations among the parts themselves. Depictive representations convey meaning via their resemblance to an object, with parts of the representation corresponding to parts of the object…. When a depictive representation is used, not only is the shape of the represented parts immediately available to appropriate processes, but so is the shape of the empty space…. Moreover, one cannot represent a shape in a depictive representation without also specifying a size and orientation…”

This quotation introduces a number of issues that need to be examined closely. One idea we can put aside is the claim that depictive representations convey meaning through their resemblance to the objects they depict. This relies on the extremely problematic notion of resemblance, which has been known to be inadequate as a basis for meaning [certainly since \Wittgenstein, 1953 #1439]. Resemblance is neither necessary nor sufficient for something to have a particular meaning or reference: Images may resemble what they do not refer to (e.g. an image of John’s twin brother does not refer to John) and they
may refer to what they do not resemble (an image of John taken through a distorting lens is an image of John even though it does not resemble him).

Despite its obvious problems, the notion of resemblance keeps surfacing in discussions of mental images, in a way that reveals how deeply the conscious experience of mental imagery contaminates conceivable theories of mental imagery. For example, (Finke 1989) begins with the observation, “People often wonder why mental images resemble the things they depict.” But the statement that images resemble things they depict is just another way of saying that the conscious experience of mental imagery is similar to the conscious experience one would have if one were to see the thing one was imagining. Consider what it would be like if images did not “resemble the things they depict”? It would be absurd if in imagining a table one had an experience that was like that of seeing a dog? Presumably this is because (a) what it means to have a mental image of a chair is that you are having an experience like that of seeing a chair, and (b) what conscious content your image has is something on which you are the final authority. You may be deceived about lots of things concerning your mental image. You may, and typically are, deceived about what sort of thing your image is (that is, what form and substance underlies it), but surely you cannot be deceived about what your mental image looks like, or what it resembles. These are not empirical facts about imagery, they are just claims about what the phrase “mental image” means.

In contrast to the vacuity of the criterion of resemblance, the proposal that images can be decomposed into ‘parts’ with the spatial relations among parts of the image in some way mapping onto the parts and the spatial relationships among the corresponding parts of the world, deserves closer scrutiny although it has not received systematic treatment in the literature. Some time ago (Sloman 1971) suggested this as a defining characteristic of analogue representations and it is clearly an important criterion. Although it needs to be spelled out in more detail, this is a reasonable proposal, but it will not yield the conclusion that images are spatial in any sense that bears on the “depiction” story. In fact, it is true of any representational system that is compositional (see section 7.1).

Another proposal mentioned in the quotation is that in depictive representations certain aspects are mandatory so that, for example if you choose to represent a particular object you cannot fail to represent its shape, orientation and size. This claim too has some truth, although the question of which aspects are mandatory, why they are mandatory, and what this tells us about the form of the representation is not so clear. It is a general property of representations that some aspects tend to be encoded (or assigned as default value) if other aspects are. Sometimes that is true by virtue of what it is that you are trying to imagine. For example, you can’t imagine a melody without also imagining each note, and therefore making a commitment as to how many notes it has. This follows from what it means to “imagine a melody,” not from the inherent nature of some particular form of representation. The same is true for other examples of imaginings. When you ask someone to imagine an familiar shape by giving its name, say the letter “B”, the person will make a commitment to such things as whether it is in upper or lower case. It seems as though you can’t imagine a B without imagining either an upper case “B” or a lower case “b”. But is this not another case of a requirement of the task to “imagine a ‘B’”? In this example, are you not being asked to describe what you would see if you were actually looking at a token of a particular printed letter? If you actually saw a token of a B you would see either a lower or an upper case letter, but not both and not neither. If someone claimed to have an image of a B that was noncommittal with respect to its case you would surely be entitled to say that the person did not have a visual image at all.

When you get to other contents of an image, the situation gets murkier because it becomes less clear what exactly the task of “imagining the letter ‘B’” entails. Does your image of the letter have to have a color or texture or shading? Must you represent the background against which you are viewing it, the direction of lighting and the shadows it casts? Must you represent it as viewed from a particular point of view? What about its stereoscopic properties; do you represent the changing parallax of its parts as you imagine moving in relation to it? Could you choose to represent any or none of these things? Most of
our visual representations, at least in memory, are noncommittal in various respects (for examples; see Pylyshyn 1978). In particular, they can be noncommittal in ways that no picture can be noncommittal. Shall we then say that they are not images? How you feel about such questions is more terminological (that is, what you are disposed to count as an image representation) than empirical. It shows the futility of assuming that mental images are just like pictures. As the graphic artist M.C. Escher once put it,

...a mental image is something completely different from a visual image, and however much one exerts oneself, one can never manage to capture the fullness of that perfection which hovers in the mind and which one thinks of, quite falsely, as something that is ‘seen’ (Escher 1960, p7).

5.2 Real versus functional space

Despite the temptation to do so, imagery theorists have been reluctant to claim that images are literally laid out in real space – that is, on a physical surface in the brain. However, because theories of imagery have had to appeal to such notions as distance, shape, size and so on, some notion of space is always presupposed. Consequently many writers who see the need for spatial properties speak of a “functional” space, with locations and other spatial properties being defined functionally (e.g., Denis & Kosslyn 1999). The example frequently cited (see the Kosslyn quotation above) is that of a matrix data structure in a computer, which can be viewed as having many of the properties of space without itself being laid out spatially in the physical machine. This is in some ways an attractive idea since it appears to allow us to claim that images have certain spatial properties without being committed to how they are implemented in the brain – so long as the implementation and its accessing operations function the way a real spatial system would function. The hard problem is to give substance to the notion of a functional space that does not reduce it to being either a summary of the data, with no explanatory mechanisms, or a model of real literal space. This problem that has been so widely misunderstood that it merits some extended discussion.

Consider first why a matrix data structure might appear to constitute a “functional space”. As typically used it seems to have two (or more) dimensions (since referencing individual cells is typically done by providing two numerical references or “coordinates”), to have distances (if we identify distance with the number of cells lying between two places), and to have empty spaces (so that it explicitly represents both where there are features and where there are no features). Graphical elements, such as points, contours, and regions can be represented by entering features into the cells at quantized coordinates. There is then a natural sense of the properties of “adjacency,” as well as of places being “between” two specified locations (as well as other simple geometrical properties of sets of features, such as being collinear, forming a triangle, and so on). Because of this, operations such as “scanning” from one feature to another, as well as of “shifting” and “rotating” patterns, can be given natural definitions (see, e.g., Funt 1980). Thus, the format of such a data structure appears to lend itself to being interpreted as “depictive” rather than “language-like” as noted in the earlier Kosslyn quote.

Notice, however, that that all the spatial notions mentioned in the previous paragraph are properties of a way of thinking about or of interpreting the data structure, they are not intrinsic properties of the matrix data structure itself. What makes cells in a matrix appear to be locations that have such properties as adjacency, betweeness, alignment, distance, and so on, is not any property of the matrix, nor even of the way that this data structure must be used. There is no sense in which any pairs of cells is special and so there is no natural sense in which some pairs of cells are “adjacent”, including a sense that derives from how they must be accessed. There are literally no constraints on the order in which cells must be accessed. We can, of course, require that the matrix be accessed in certain ways, and when we model imagery we typically do stipulate that certain pairs of cells be considered “adjacent” and that in accessing any pair of cells in a serial fashion, certain other cells (the ones we designate as being “between” the pair) must be visited first and in a certain order (which we call “scanning”). But it is critical to the interpretation of a computational process as a model of mental imagery that we be clear as
to why such constraints hold. If our model of imagery assumed a literal physical surface, then the
reason would be clear: physical laws require that movement over a surface follow a certain pattern, such
as that the time it takes to get from one place to another is the ratio of the distance traversed to the speed
of movement. But in a matrix no such intrinsic constraint exists. Such a constraint must be stipulated as
an extrinsic constraint (along with many other constraints, such as those that govern the invariance of
adjacency, betweenness, or collinearity, with transformations of scale, orientation, and translation). The
spatiality of a matrix, or of any other form of “functional space”, must be stipulated or assumed over and
above any intrinsic property of the format of the representation. The crucial fact about extrinsic
constraints is that such constraints are independently stipulated, and so could be applied equally to any
form of representation, including a model of imagery that used symbolic expressions or structured
descriptions. So far as the notion of functional space is concerned, there is nothing to prevent us from
modeling the content of an image as a set of sentence-like expressions in a language of thought. We
could then stipulate that in order to go from examining one place (referred to, say, by a unique name) to
examining another place (also referred to by a name) you must pass through (or apply an operation to)
the places whose names are located between the two names on some list. You might object that this sort
of model is ad hoc. It is. But no more ad hoc than when the constraints are applied to a matrix
formalism. Notice, moreover, that both become completely principled if they are taken to be
simulations of a real spatial display.

You might wonder why the matrix feels more natural than other ways of simulating space. The
answer may be that a matrix offers a natural model of space because we are used to thinking of and
displaying matrices as two-dimensional tables (complete with empty cells) and of viewing the cells as
being referenced by names that we think of as pairs of coordinates. We thus find it easy to switch
back and forth between the data-structure view and the (physical) table view. Because of this, it is
natural to interpret a matrix as a model of real space and therefore it is easy to make the slip between
thinking of it on one hand as merely a “functional space” and thinking of it, on the other hand, as a
stand-in for (or a simulation of) real space – a slip we encounter over and over in theorizing about the
nature of mental imagery. As a simulation of real space it is unproblematic so far as the sorts of
problems discussed here are concerned. But we must recognize that in this case we are assuming that
images are written on a literal spatial medium, which we happen to be simulating by a matrix (for
reasons of convenience). In fact in (Kosslyn et al. 1979) this view was made explicit when the authors
invoke what they call the “cathode-ray tube model”. In that case it is the literal space that has the
explanatory force, notwithstanding the fact that, as a practical matter, it is being simulated on a digital
computer.

The point is that there is no such thing as a “functional space” apart from the set of extrinsic
stipulations or constraints we choose to impose on such things as how symbolic names (e.g., matrix
coordinates) map onto places in a physical display and how distances and geometrical predicates are to
be interpreted over the data structure. What we have, rather, is one of two things: either a real physical
space, with its (approximately) Euclidean properties, or a symbolic model of such a space. Anything
else is merely metaphorical and not explanatory. It allows one to think of an image as spatial without the
attending disadvantages of having made an untenable assumption about the architecture of mental
imagery.

The real scientific question is not how we can model space in a theory of mental imagery. Rather, it
is whether there is any sense in which the architecture of mental imagery incorporates the geometry of
real space. Only after we have answered this empirical question can we know whether one should
model properties of space in modeling imagery. My purpose in belaboring the distinction between
intrinsic and extrinsic constraints, and what is being presupposed when we talk of “functional space,” is
simply to set the stage for the real issues, which are empirical. I have already described some of the
relevant empirical findings in connection with the mental scanning and have suggested that the same is
likely to be true for other findings that imply that images have metrical properties. The cognitive
penetrability of such phenomena suggests that the mind does not work as though the imagery architecture imposes constraints like those you would expect of a real spatial display. It appears that we are not required to scan through adjacent places in getting from one place to another in an image – we can get there is quickly or as slowly as we wish, with or without visiting intermediate filled or empty places (assuming that visiting empty places is even possible – see section 0).

5.3 Projected mental images: Inheriting spatial properties from real space

In most imagery studies subjects are asked to imagine something while looking at a scene; thus, at least in some phenomenological sense, superimposing or projecting an image onto the perceived world. Yet it has been amply demonstrated (O'Regan & Lévy-Schoen 1983) that true superposition of visual percepts does not occur when visual displays are presented in sequence, or across saccades. So what happens when a mental image (whether constructed or derived from memory) is superimposed over a scene? In many of these cases (e.g., Farah 1989; Hayes 1973; Podgorny & Shepard 1978) a plausible answer is that one allocates attention to the scene according to a pattern that corresponds roughly to the projected image. Alternatively, and more plausibly, one simply thinks of imagined objects as being located at places actually occupied by certain perceived ones. Thinking that something is at a certain location need not entail projecting an imagined shape onto some background. It might require nothing more that allocating attention to a particular object in a scene and thinking of that object as having a certain property. It is no more than thinking “this (e.g., referring to a bit of texture) is where I imagine feature F to be located”.

The capacity for this sort of “demonstrative reference” has been investigated extensively and discussed by (Pylyshyn 2000, 2001).

Consider, for example, the study reported by (Podgorny & Shepard 1978). In their vision control condition, experimenters asked subjects to indicate as fast as possible whether a dot appeared on or beside a simple figure, such as the letter F. The pattern of reaction times was recorded (e.g., times were found to be shorter when the dot was ON the figure than OFF and shorter when it was at a vertex rather than mid-stroke, etc). This pattern was then compared with the pattern obtained when the figure was merely imagined to be on the grid while a real dot was presented at corresponding locations (as shown in Figure 4). In the image condition, the pattern of reaction times obtained was very similar to the one obtained from the corresponding real display. This was interpreted as showing that in both vision and projected imagery, subjects perceived a similar visual pattern. But a more parsimonious account is that in imagining the figure in this task, subjects merely attended to the rows and columns in which the imagined figure would have appeared. We know that people can indeed direct their attention to several objects such as rows or columns or cells in a display, or even to conform their attention to a particular shape. Focusing attention in this way is all that is needed in order to generate the observed pattern of reaction times. In fact using displays similar to those used in the (Podgorny & Shepard 1978) study, but examining the threshold for detecting spots of light, (Farah 1989) showed that the instruction to simply attend to certain letter-shaped regions was more effective in enhancing detection in those regions than instructions to superimpose an image over the region.

---

**Figure 4.** Observers were shown a figure (display 1) which they then had to retain as an image and to indicate whether the dot (display 2) occurred on or off the imagined figure. The pattern of reaction times was found to be similar to that observed when the figure was actually present. (based on Podgorny & Shepard 1978).
A similar story applies to other tasks that involve responding to image properties when images are superimposed over a perceived scene (e.g., Hayes 1973). If, for example, you imagine the map used to study mental scanning (shown in Figure 1) superimposed over one of the walls in the room, you can use the visual features of the wall to anchor various objects in the imagined map. You can think a thought which might be paraphrased as “the church is located where this (speck) is on the wall, the beach is beside that (corner) ...”, where each of the locative terms this and that picks out an object in the visual field and binds it to terms in the thought. Once the appropriate items are bound, “scanning the image” is accomplished by scanning between the selected items in the actual visual display. Thus the increase in the time it takes to scan between items that are further apart on the imagined map is easily explained, since it involves scanning greater distances in the real scene. In general, such cases of superposition allow many of the spatial properties of the real scene (e.g., properties expressed by Euclidean and metrical axioms) to be inherited by the combined image-percept. For example, if image features A, B and C are imagined to be collinear and they are bound to three actual collinear features in some visible scene, then the fact that feature B is between A and C can be visually read off the perceived scene.

The mechanism for indexing imagined objects to visual features, called visual indexes, has been described extensively elsewhere (e.g., Pylyshyn 1994b; Pylyshyn 1998, 2000, 2001). Such a binding is an example of the use of indexicals in image representations, and illustrates one way in which symbols underlying imagery can be what (Barsalou 1999a) calls perceptual symbols. This also appears to be what (Glenberg & Robertson 2000) have in mind when they speak of image symbols as being grounded or embodied (this sort of grounding is discussed extensively in Pylyshyn 2001).

Note that this vision-plus-indexes story is far from being equivalent to superimposing an image over the scene, because it assumes no pictorial properties of the “superimposed” image, only the binding of imagined objects/locations to real perceived ones. Moreover, because the relevant information involves only sparse spatial locations and not other detailed visual properties, the memory demand is minimal, so this process might conceivably be carried out even for short time after the scene has been removed, as is the case when subjects are asked to close their eyes when they recall the image (there is evidence that short term recall of low-level iconic information can persist for several minutes; see Ishai & Sagi 1995). Thus this sort of story might provide another mechanism to explain phenomena such as mental scanning, even when scanning is carried out with eyes closed.

5.4 Visuomotor interaction with images

Another reason to think of mental images as being spatial is that they appear to connect with spatial aspects of the motor system in a way that is similar to how visually perceived space connects with the motor system. For example, it seems that you can imagine a scene and point to some object in the imagined scene, which suggests that the image is in some sense spatial. This apparent coordination with movements is surely one major reason to think of images as spatial since motor control commands must be issued in spatial coordinates. If we could move our eyes to a place in an image or point to a feature of the image we would be exhibiting some aspect of the image’s spatiality.

This raises the empirical question of whether images engage the motor system the way vision does. In a series of ingenious experiments, (Finke 1979) showed that adaptation to displacing prisms could be obtained using only the imagined location of the observer’s hand as feedback. These studies are of special interest because they illustrate the way in which projected images can work like real percepts. Finke asked subjects to imagine that their (hidden) hand was in a certain specified location. The sequence of locations where he asked them to imagine their hand actually corresponded to the errors of displacement made by another subject who had worn displacing prisms. Finke found that both the pattern of adaptation and the pattern of after-effects shown by the subjects who were asked to imagine the location of their hand was similar to that exhibited by subjects who actually wore displacing prisms and so could see the displaced location of their hand.
It is known that adaptation can occur in response to cognitive factors, and indeed even to verbally presented error information (Kelso, Cook, Olson, & Epstein 1975; Uhlarik 1973), though in that case the adaptation occurs more slowly and transfers completely to the nonadapted hand. Yet Finke found that in the case of imagined hand position, the adaptation, though significantly lower in magnitude, followed the pattern observed with the usual visual feedback of hand position. Moreover, when subjects were told that their hand was actually not where they imagined it to be, the adaptation effects was nonetheless governed by the imagined location, rather than where they were told their hand was, and followed the same pattern as that observed with both imagery and actual visually presented error information.\[9\] It thus seems that the visuomotor system may be involved when adapting to imagined hand positions. The question is: What is the nature of this involvement?

Exactly what causes prism adaptation has been the subject of some debate (Howard 1982), but it is generally accepted that an important factor is the discrepancy between the seen position and the felt position of the hand (or the discrepancy between visual and kinesthetic or proprioceptive location information). Significantly, such discordance does not require that the visual system recover any visual property of the hand other than its location. Indeed, in some studies of adaptation, subjects viewed a point source of light attached to their hand rather than the hand itself (Mather & Lackner 1977) with little difference in adaptation. But exactly where the subject attends is important (Canon 1970, 1971). In some cases even an immobile hand can elicit adaptation providing the subject visually attends to it (Mather & Lackner 1981). In Finke’s experiments, subjects focused their gaze towards a particular location, where they were, in effect, told to pretend (incorrectly) that their hand was located, thus focusing attention on the discordance between this imagined location of their hand and their kinesthetic and proprioceptive sense of the position of their arm.\[10\] Thus the imagery condition in these studies provides all that is needed for adaptation – without requiring any assumptions about the nature of imagery.

It seems that there are a number of imagery-motor phenomena that depend only on orienting one’s gaze or one’s focal attention to certain perceived locations. The Finke study of adaptation of reaching is a plausible example of this sort of phenomenon, as is the (Tlauka & McKenna 1998) study of S-R compatibility for manually responding to information in images. None of these results require that imagery feed into the visuomotor system, let alone that images be spatial or depictive. Indeed, these cases involve actual visual perception of location (that is, there really are visible elements at the relevant locations in real space that are being attended). Since the adaptation phenomena (as well as the S-R compatibility phenomena) require only location information and no pictorial information (e.g., shape, color, etc), they do not in any way implicate the “depictive” character of mental images.

When we look at cases in which images are not projected onto a perceived scene (say they are imagined in the dark or with eyes closed), or in which more than just the imagined location of an object is relevant to the motor action, we find that images do not interact with the perceptual motor system in a way that is characteristic of visual interaction with that system. To show this we need to look at certain signature properties of the visual control of movements, rather than at cases where the control may actually be mediated only by spatial attention. One clear example of strictly visual control of motor action is smooth pursuit. People can track the motion of slowly moving objects with a characteristic sort of eye movement called smooth pursuit. There are also reports that under certain circumstances people can track the voluntary (and perhaps even involuntary) movement of their hand in the dark (Mather & Lackner 1980). They can also track the motion of objects that are partially hidden from view (Steinbach 1976), and even of induced (apparent) motion of a point produced by a moving frame surrounding the point (Wyatt & Pola 1979). In other words they can smoothly pursue inputs generated by the early vision system. Yet what people cannot do is smoothly pursue the movement of imagined objects. In fact it appears to be impossible to voluntarily initiate smooth pursuit tracking without a moving stimulus (Kowler 1990).
Another example of a characteristic visually guided control is reaching to grasp a visible object. Although we can reach out to grasp imagined objects, when we do so we are essentially pantomiming a reaching movement, rather than engaging the visuomotor system. Visually guided reaching exhibits certain quite specific trajectory properties not shared by pantomimed reaching (Goodale, Jacobson & Keillor 1994). For example, the time and magnitude of peak velocity, the maximum height of the hand, and the maximum grip aperture, are all significantly different when reaching to imagined that to perceived objects. Reaching and grasping gestures towards imagined objects exhibit the distinctive pattern that is observed when subjects are asked to pantomime a reaching and grasping motion. There is considerable evidence that the visuomotor system is itself encapsulated (Milner & Goodale 1995) and, like the visual system, is able to respond only to information arriving from the eyes, which often includes visual information that is not available to consciousness. As with the visual system, only certain limited kinds of modulations of its characteristic behavior can be imposed by cognition. When we examine signature properties of the encapsulated visuomotor system, we find that mental images do not engage it the way that visual inputs do.

6 Are images “seen” by the visual system?

One of the most actively pursued questions in contemporary imagery research has been the question of whether mental imagery uses the visual system. Intuitively the idea that imagery involves vision is extremely appealing since the experience of imagery is phenomenally very like the experience of seeing (indeed there have been (disputed) claims that when real perception is faint because of impoverished stimuli, vision and imagery may be indistinguishable, Perky 1910). But, from the perspective of the present thesis, there is a more interesting reason for asking whether the visual system is involved in examining images. If vision is used to interpret mental images it might support the idea that images are things that can be seen, thus lending credence to the intuitive view of images as pictorial (or depictive). The question of the overlap between imagery and vision has been investigated with particular zeal within the cognitive neuroscience community. We shall examine the neuroscience findings in section 0. For the present I want to consider some of the psychological evidence that suggests an overlap between imagery and visual perception, and to ask whether this evidence supports the view that images are pictorial objects that are “seen” by the visual system.

6.1 The experience of seeing and of imagining

It may well be that the most persuasive reason for believing that mental imagery involves the visual system is the subjective one: Mental imagery is accompanied by a subjective experience that is very similar to that of seeing. As we remarked at the beginning of this article, this sort of phenomenal experience is very difficult to ignore. Yet it is quite possible that both vision and imagery lead to the same kind of experience because the same symbolic, rather than pictorial, form of representation, underwrites them both. The experience might correspond to a fairly high level of the analysis of the visual information, say at the point where the stimulus is recognized as something familiar (Crick & Koch 1995; Stoerig 1996, suggest that the locus of our awareness occurs higher than primary visual cortex). In that case even though imagery and vision shared common mechanisms and forms of representation, one could not infer that the form was depictive or pictorial. At most one might conclude, as does (Barsalou 1999b), that the form of representation underlying images, while symbolic in nature, is also modality-specific inasmuch as it consists of some subset of the neural activity that is associated with the corresponding visual perception. This alternative is compatible with our proposal (next section) that the representation underlying vision and visual imagery may use the same modality-specific symbolic vocabulary.

6.2 Interference between imaging and visual perception

One of the earliest findings that persuaded people that images involve the visual system was that the task of examining images could be disrupted by a subsidiary visual or spatial task. For example,
(Brooks 1968) showed that reporting spatial properties from images is more susceptible to interference when the response must be given by a spatial method (e.g. pointing) than by a verbal one (that is, speaking). When subjects were asked to describe the shape of the letter F by providing a list of the right and left turns one would have to take in traveling around its periphery, their performance was worse when they had to point to the left or right (or to left- and right-pointing arrows) than when they had to say the words “left” and “right”. (Segal & Fusella 1969, 1970) subsequently confirmed the greater interference between perception and imagery in various same-modality tasks and also showed that both sensitivity and response bias measures (derived from Signal Detection Theory) were affected. Segal and Fusella concluded that “imagery functions as an internal signal which is confused with the external signal” (p. 458). This may be the correct conclusion to draw. But this does not show that either of the inputs is pictorial. All that it implies is that the same type of representational contents are involved, or to put it another way, that the same concepts are deployed. For the sake of argument, think of the representations in these studies as being in a common language of thought: What, in that case, do the representations of visual patterns have in common with mental images of visual patterns? It is the fact they are both about the appearance of visual patterns. Like sentences about visual appearances, they all involve the use of concepts such as “bright,” “red,” “right angle,” “parallel to” and so on. It is not surprising that two responses requiring the same modality-specific conceptual vocabulary would interfere. Thus it may be that visual percepts and visual images interact because both consist of symbolic representations that use some of the same proprietary spatial or modality-specific vocabulary. That the linguistic output in the Brooks study is not as disruptive as pointing may simply show that spatial concepts are not relevant to articulating the words “left” or “right” once they have been selected for uttering, whereas these concepts are relevant to issuing the motor commands to move left or right.

6.3 Visual illusions induced by superimposing mental images

Other studies that are cited in support of the view that images are interpreted by the visual system are studies showing that projecting images of certain patterns onto displays creates some of the well-known illusions, such as the Müller-Lyer illusion, the Pogendorf illusion or the Herring illusion, or even the remarkable McCollough effect. For example, (Bernbaum & Chung 1981) showed subjects displays such as those in the top part of Figure 5. Subjects were asked to imagine lines connecting the endpoints of the visible line to either the outside or the inside pairs of dots in this display (when the endpoints are connected to the inside pair of dots they produce outward-pointing arrows and when they are connected to the outside pair of dots they produce inward pointing arrows, as in the original Müller-Lyer illusion). Bernbaum & Chung (also Ohkuma 1986) found that imagining the arrows also produced the illusion, with the inward pointing arrows leading to the perception of a longer line than the outward pointing arrows. For the sake of argument let us take these results as valid, notwithstanding the obvious susceptibility of such findings to experimenter demand effects – see note 11.

```
  * *

  * * * * * * * * *

  * *

  * * * * * * * * *

  *

  *
```

* Figure 5 Figures used to induce the Müller-Lyer illusion from images. Imagine the end points being connected to the inner or the outer pairs of dots in the top figure (Bernbaum & Chung 1981) or selectively look at the inward or outward arrows in the bottom figure (based on Goryo, Robinson & Wilson 1984).
Before one can interpret such findings one needs to understand why the illusion occurs in the visual case. Explanations for the Müller-Lyer and similar illusions tend to fall into one of two categories. They either appeal to the detailed shapes of contours involved and to the assumption that these shapes lead to erroneous interpretations of the pattern in terms of 3D shapes, or they appeal to some general characteristics of the 2D envelope created by the display and the consequent distribution of attention or eye movements. The former type of explanation, which include the “inappropriate constancy scaling” theory due to (Gregory 1965), have not fared well in general since the illusion can be produced by a wide variety of types of line-endings (see the review in Nijhawan 1991). The latter type of explanation, which attributes the illusion to the way attention is allocated and to mechanisms involved in preparing eye-movement, have been more successful. For example, one theory (Virsu 1971) appeals to the tendency to move ones eyes to the center of gravity of a figure. The involvement of eye movements in the Müller-Lyer illusion has also been confirmed by (Bolles 1969; Coren 1986; Festinger, White & Allyn 1968; Hoenig 1972; Virsu 1971). Another example of the envelope type of theory is the framing theory (Brigell, Uhlarik & Goldhorn 1977; Davies & Spencer 1977) which uses the ratio of overall figure length to shaft length as predictor. Such envelope-based theories have generally fared better than shape-based theories, not only on the Müller-Lyer illusion, but also for most cases in which there are context effects on judgments of linear extent. What is important about this for our purposes is that these explanations do not appeal to pattern-perception mechanisms and therefore are compatible with attention-based explanations of the illusions.

Further evidence that attention can play a central role in these illusions (as well as in other visual illusions induced by mental imagery, e.g., Wallace 1984a, 1984b) comes from studies that actually manipulate attention focus. For example, it has been shown (Goryo et al. 1984) that if both sets of inducing elements (the outward and inward arrowheads) were present (as in the bottom part of Figure 5), observers could selectively attend to one or the other and obtain the illusion appropriate to the one to which they attended. This is very similar to the effect demonstrated by (Bernbaum & Chung 1981) but without requiring that any image be superimposed on the line. (Coren & Porac 1983) also confirmed that attention alone could create, eliminate or even reverse the Müller-Lyer illusion. Attention-mediation was also shown explicitly in the case of an ambiguous motion percept (Watanabe & Shimojo 1998). This is in keeping with the evidence we considered in section 5.3 showing that in many cases in which mental images are interpreted as having visual effects, the effect can be explained by appeal to the attention-focusing role that imagery plays in connection with visual perception. Finally, the relevance of the imagined induction of the Müller-Lyer and similar illusions to the picture theory is further cast into doubt when one recognizes that such illusions, like many other imagery-based phenomena, also appears in congenitally blind people (Patterson & Deffenbacher 1972).

6.4 Imagined versus perceived motion

(Gilden, Blake & Hurst 1995) used visual motion adaptation to study whether the visual system is involved in imagery. This study is of special interest to us since motion adaptation is known to be retinotopic, and therefore occurs in the early visual system. When a region of the visual field receives extensive motion stimulation, an object presented in that region is seen to move in the opposite direction to the inducing movement (this is called the “waterfall illusion”) and a moving object is seen as moving more slowly. Gilden, et al., designed their study with the intention of showing that the motion of an imagined object is affected by the aftereffect of a moving field. They had subjects gaze for 150 seconds at a square window on a screen containing a uniformly moving random texture. Then they showed subjects a point moving towards that window and disappearing behind what appeared to be an opaque surface, and they asked subjects to imagine the point continuing to move across the previously stimulated region and to report when the point would emerge at the other side of the surface. Gilden et al. did find an effect of motion adaptation on imagined motion, but it was not exactly the effect they had expected. They found that when the point was imagined as moving in the same direction as that of the inducing motion field (that is, against the motion aftereffect) it appeared to slow down (it took longer to
reach the other side of the region). However, when the point was imagined as moving in the opposite direction to the inducing motion field (that is, in the same direction as the motion aftereffect), the point appeared to speed up (it reached the other side in a shorter time). The latter effect is not what happens with real moving points. In visual motion adaptation, motion appears to slow down no matter which direction the inducing motion field moves, presumably because all motion sensitive receptors have been fatigued. But, as Gilden et al. recognized, the effect they observed is exactly what one would expect if, rather than imagining the point moving continuously across the screen, subjects imagined the point as being located at a series of static locations along the imagined path. This suggests a quite different mechanism underlying imagined motion when the latter is generated as the extrapolation of perceived motion. We know that people are very good at computing time-to-contact of a uniformly moving object at a specified location (e.g., DeLucia & Liddell 1998). What may be going on in imagined motion is that people may simply pick out one or more marked places (e.g., elements of texture) along the path, using the visual indexing mechanism discussed earlier, and then compute the time-to-contact for each of these places.

We explicitly tested this idea (Pylyshyn & Cohen 1999) by asking subjects to extrapolate the motion of a small square, which disappeared by occlusion behind an apparent opaque surface. They were asked to imagine the smooth motion of the square in a dark room. At some unpredictable time in the course of this motion the square would reappear, as though coming out through a crack in the opaque surface, and then receded back through another crack, and subjects would indicate whether this reappearance occurred earlier or later than when their imagined square reached that crack. This task was carried out in several different conditions. In one condition the location of the “cracks” where the square would appear and disappear was unknown (that is, the cracks were invisible). In another condition the location at which the square was to appear was known in advance: it was indicated by a small rectangular figure that served as a “window” through which, at the appropriate time, subjects would briefly view the square that was moving behind the surface (the way the squares appeared and disappeared in the window condition was identical to that in the no-window condition except that the outline of the window was not visible in the latter case). And finally in one set of conditions the imagined square moved through total darkness whereas in the other set of conditions the path was marked by a sparse set of dots that could be used as reference points to compute time-to-contact. As expected, the ability to estimate where the imagined square was a various times (measured in terms of decision time) was significantly improved when the location was specified in advance and also when there were visible markers along the path of imagined motion.

Both of these findings confirm the suggestion that what subjects are doing when they report “imagining the smooth motion of a square” is selecting places at which to compute time-to-contact and are merely thinking that the imaginary moving square is at those places at the estimated times. According to this view, subjects are thinking the thought “now it is here” repeatedly for different visible objects (picked out by the visual indexing mechanism mentioned earlier), and synchronized to the independently computed arrival times. This way of describing what is happening requires neither the assumption that the visual system is involved (other than the attentional indexing mechanism) nor does it need to assume that an imagined square is actually moving through some mental space and occupying each successive position along a real spatial path. Indeed there is no need to posit any sort of space except the visible one that serves as input to the time-to-contact computation.

6.5 Extracting novel information from images: Visual (re)perception or inference?

One of the alleged purposes of mental images is that they can be examined in order to visually discover new properties or new interpretations or reconstruals. It would therefore seem important to ask whether there is any evidence for such visual reconstruals. This question turns out to be more difficult to answer than one might have expected, for it is clear that by examining images one can draw some conclusions that were not explicitly present in, say, the verbal description under which it was imagined. So if I ask you to imagine a square and then to imagine drawing in both diagonals, it does not seem
surprising if you can tell that the diagonals cross or that they form an “X” shape. Since this inference is very simple it does not seem prima facie to qualify as an example showing that images are interpreted visually. On the other hand, suppose I ask you to imagine two parallelograms, one directly above the other on a page, and then to connect each vertex of the top one to the corresponding vertex of the bottom one. What do you see? As you keep watching, what happens in your image? When presented visually, this figure leads to certain phenomena that do not appear in mental imagery. The signature properties of spontaneous perception of certain line drawings as depicting three-dimensional objects and spontaneous reversals of ambiguous figures do not appear in this mental image (which happens to be a Necker Cube).

But what counts, in general, as a visual interpretation as opposed to an inference? I doubt that this question can be answered without a sharper sense of what is meant by the term “visual”. Since the everyday (pretheoretical) sense of the notion of “vision” clearly involves most, if not all, of cognition, the question of the involvement of vision in reconstruals cannot be pursued without a more restricted sense of what counts as a visual phenomenon. Clearly, deciding whether two crossed lines form an “X” is not one of these phenomena, nor is judging that when a D is rotated counterclockwise by 90 degrees and placed on top of a J the result looks like an umbrella. You don’t need to use the early visual system in deciding that. All you need is an elementary inference based on what makes something look like an umbrella (e.g., it has an upwardly convex curved top attached below to a central vertical stroke – with or without a curved handle at the bottom). Thus these the sorts of examples (which were used in Finke, Pinker & Farah 1989), cannot decide the question of whether images are visually (re)interpreted. (Hochberg 1968) suggested a few signature properties of vision, including spontaneous interpretation of certain line drawings as 3-dimensional objects, spontaneous reversal of ambiguous shapes, and spontaneous interpretation of certain sequences as apparent motion. Although these criteria cannot always be used to decide whether a particular interpretation is visual, they do indicate the sort of constructs that appear to be constitutive of early vision. For the time being all we can do is ask whether the candidate reconstrual meets this sort of intuitive condition. A more thorough analysis would attempt to look at converging evidence concerning the level of the visual system implicated.

The clearest evidence I am aware of that bears on that question of whether images are subject to visual reconstruals is provided by studies carried out by Peter Slezak (Slezak 1991, 1992, 1995). Slezak asked subjects to memorize pictures such as those in Figure 6. Then he asked them to rotate the images clockwise by 90 degrees and to report what they looked like. None of his subjects was able to report the appearance that they could easily report by rotating the actual pictures. The problem was not with their recall or even their ability to rotate the simple images; it was with their ability to recognize the rotated image in their mind’s eye. We know that because subjects were able to draw the figures from memory and when they rotated those they did see the other construals. What is special about these examples is that the resultant appearance is so obvious – it comes as an “aha!” experience when carried out by real rotation. Unlike the figures used by (Finke et al. 1989), these shapes were not familiar and their appearance after the transformation (by rotation) could not be easily inferred from their representation.

Figure 6. Orientation-dependent figures used by(Slezak 1991). To try these out, memorize the shape of one or more of the figures, then close your eyes and imagine them rotated clockwise by 90 degrees (or even do it while viewing
A related question that can be asked is whether images can be ambiguous, since this also concerns the question of whether images can be visually reinterpreted. This case, however, presents some additional methodological problems. Not all ambiguities contained in pictures are visual ambiguities, just as not all reinterpretations are visual reconstruals. For example, the sorts of visual puns embodied in some cartoons or most characteristically in so-called “droodles” (see http://www.droodles.com for examples) rely on ambiguities, but clearly not on ones that are based in part on different visual organizations being produced by early visual processes. By contrast, the reversal of figures such as the classical Necker Cube is at least in part the result of a reorganization that takes place in early vision. Do such reorganizations occur with visual images? In order to answer that questions we would have to control for certain obvious alternative explanations of any reports of apparent reinterpretations. For example, if a mental image appeared to reverse, it might be because the observer knew of the two possible interpretations and simply replaced one of its interpretations with the other. This is the view that many writers have taken in the past (Casey 1976; Fodor 1981).

(Chambers & Reisberg 1985) were the first to put the question of possible ambiguous mental images to an empirical test. They reported that no reversals or reinterpretations of any kind took place with mental images. Since that study was reported there have been a series of studies and arguments concerning whether images could be visually (re)interpreted. (Reisberg & Chambers 1991; Reisberg & Morris 1985) used a variety of standard reversal figures and confirmed the Chambers and Reisberg finding that mental images of these figures could not reverse. (Finke et al. 1989) took issue with these findings, citing their own experiments involving operations over images (mentioned above), but as I suggested, it is dubious that the reinterpretation of their simple familiar figures should be counted as a visual reinterpretation. Even if these could be considered visual interpretations, the more serious problem is to explain why clear cases of visual interpretations, such as those studied by Chambers and Reisberg, do not occur with images.

A more direct and extensive exploration of whether mental images can be ambiguous was undertaken by (Peterson 1993; Peterson, Kihlstrom, Rose, & Glisky 1992) who argued that certain kinds of reconstruals of mental images do take place. Peterson first distinguished different types of image reinterpretations. In particular, she distinguished what she calls reference-frame realignments (in which one or more global directions are reassigned in the image, as in the Necker cube or rabbit-duck ambiguous figures) from what she calls reconstruals (in which reinterpreting the figure involves assigning new meaning to its parts, as in the wife/mother-in-law or snail/elephant reversing figures). We will refer to the latter as part-based reconstruals to differentiate them from other kinds of reconstruals (since their defining characteristic is that their parts take on a different meaning). A third type, figure-ground reversal (as in the Rubin vases), was acknowledged to occur rarely if ever with mental images (a finding that was also confirmed by Slezak 1995, using quite different displays). Among her findings, Peterson showed that reference-frame realignments do not occur in mental images unless they are cued by either explicit hints or implicit demonstration figures, whereas some part-based reconstruals occurred with 30% to 65% of the subjects.

Recall that our primary concern is not with whether any reinterpretations occur with mental images. The possibility of some reinterpretation depends upon what information or content-cues are contained in the image, which is orthogonal to question of the mechanisms used in processing it. What we are concerned with is whether the format of images is such that their interpretation and/or reinterpretation involves the specifically visual (that is, the early vision) system as opposed to the general inference system. The crucial question, therefore, is how Peterson’s findings on reinterpreting mental images compares with the reinterpretations observed with ambiguous visual stimuli. The answer appears to be that even when reinterpretations do occur with mental images, they are qualitatively different from those that occur with visual stimuli. For example, (Peterson 1993) showed that whereas reference-frame reversals are dominant in vision they are rare in mental imagery while the converse is true for part-based
reconstruals. Also the particular reconstruals observed with mental images tend to be different from those observed with the corresponding visual stimuli. Visual reconstruals tend to fall into major binary categories – in the case of the figures used by Peterson et al., these are the duck-rabbit, or the snail-elephant categories. On the other hand, in the imagery case subjects provided a large number of other interpretations (which, at least to this observer, did not seem to be clear cases of distinctly different appearances – certainly not as clear as the cases of the Necker Cube reversal or even the reconstruals observable when the shapes in Figure 6 are rotated). The number of subjects showing part-based reconstruals with mental images dropped by half when only the particular interpretations observed in the visual case are counted. Reinterpretation of mental images is also highly sensitive to hints and strategies, whereas there is reason to doubt that the early stages of vision are sensitive to such cognitive influences (Pylyshyn 1999).

The reason for these differences between imagery and vision is not clear, but they add credence to the suggestion that what is going on in the mental image reconstruals is not a perceptual (re)interpretation of a generated picture, but something else. Perhaps what is going on is inference and memory retrieval based on shape properties – the sort of process that goes on in the decision stage of vision, after early vision has generated shape-descriptions. This is the stage at which beliefs about the perceived world are established so we expect it to depend on inferences from prior knowledge and expectations, like all other cases of belief fixation. It seems quite likely that parts of the highly ambiguous (though not clearly bistable) figures used by Peterson et al., might serve as cues for inferring or guessing at the identity of the whole figure (for illustrations of these figures; see Peterson 1993). Alternatively, as suggested earlier, several possible forms might be computed by early vision (while the figures were viewed) and stored, and then during the image-recall phase a selection might made from among them based on a search for meaningful familiar shapes in long-term memory. While in some sense all of these are reinterpretations of the mental images, they do not all qualify as the sort of visual “reconstruals” of images that show that mental images are pictorial entities whose distinct perceptual organization (and reorganization) is determined by the early vision system. Indeed they seem more like the kind of interpretations one gets from Rorschach inkblots.

7 Can evidence from neuroscience settle the question?

7.1 Searching for the “mind’s eye” and the “image” in the brain

The neuroscience research concerned with mental imagery has been devoted primarily to attempting to show that the early visual system is involved in mental imagery. The involvement of visual mechanisms in mental imagery is of interest to the picture theorists primarily because of the possibility that the particular role played by the visual system in processing mental images will vindicate a version of the picture theory, by showing that imagery does indeed make use of a special sort of spatial display (this is explicitly the claim in Kosslyn 1994). The question that naturally arises is whether we can make a case for this view by obtaining evidence concerning which areas of the brain are involved in mental imagery and in visual perception. Before examining the evidence it is worth reiterating that the question of the involvement of the visual system and the question of the form of mental images are largely independent questions. It is logically possible for the visual system to be involved in both vision and mental imagery and yet in neither case generate picture-like depictive representations. Similarly it is possible for representations to be topographically organized in some way and yet have nothing to do with visual perception, nor with any depictive character of the representation.

In a certain sense the physical instantiation of any adequate cognitive representation must be topographically organized. Fodor and I (Fodor & Pylyshyn 1988) have argued that any form of representation that is adequate as a basis for cognition must be compositional, in the sense that the content of a complex representation must derive from the content of its parts and the rules by which the complex is put together (that is, the way the meaning of sentences is compositional and depends on the meaning of its parts together with how they are syntactically put together). The physical instantiation of
any representation that meets the requirement of compositionality will itself be compositional (Pylyshyn 1984, pp54-69; Pylyshyn 1991b). In the case of symbolic representations, parts of expressions are mapped recursively onto parts of physical states and syntactic relations are mapped onto physical relations. As a result, there is a very real sense in which the criteria in the Kosslyn quotation at the beginning of section 5.1 are met by any adequately expressive physical symbol system, not just a depictive one. Note that in a digital computer, data-structure representations are compositional and topographically distributed and yet are generally not thought to be depictive, whereas when they are supposed to be depictive, as when they encode images (say as JPEG or TIFF files), their topographical distribution generally does not mirror the physical layout of the picture. Consequently, the question of the spatial distribution of images, the question of whether they are depictive, and the question of whether they are connected with vision are logically independent questions.

Notwithstanding the logical independence of the question of the format of images and the question of the involvement of the visual system, the following line of rerasoning continues to hold sway in the neuroscience literature (Kosslyn, Pascual-Leone, Felican, Camposano, Keenan, Thompson, Ganis, Sukel, & Alpert 1999a; Kosslyn, Thompson, Kim, & Alpert 1995). Primary visual cortex (Area 17) appears to be organized retinotopically (at least in monkey brain). So if Area 17 is activated when subjects generate mental images, then it must be that higher-level cortical centers are generating retinotopically-organized activity in the visual system. In other words, during mental imagery a spatial pattern of activity is generated in the same parts of the visual system where such activity occurs in vision. From which it follows (so the argument goes) that a spatial or “depictive” form of activity is laid out in the cortex, much as it is on the retina, and this activity is then interpreted (or “perceived”) by the visual system. This line of reasoning is very much in keeping with the views held by proponents of the subjectively satisfying picture theory. It is no surprise then that those who hold the picture view welcome any evidence of the involvement of early vision in mental imagery.

With this as background, we can ask whether there is evidence for the involvement of early, topographically organized areas of vision cortex in mental imagery and, if so, whether the involvement is of the right kind. Some evidence has been reported showing that mental imagery involves activity in areas of striate cortex associated with vision. Most of this evidence comes from studies using neural imaging to monitor regional cerebral blood flow. While some neural imaging studies report activity in topographically organized cortical areas (Kosslyn et al. 1999a; Kosslyn et al. 1995), most have reported that only later visual areas, the so-called visual association areas, are active in mental imagery (Charlot, Tzourio, Zilbovicius, Mazoyer, & Denis 1992; Cocude, Mellet & Denis 1999; D'Esposito, Detre, Aguirre, Stallcup, Alsop, Tippet, & Farah 1997; Fletcher, Shallice, Frith, Frackowiak, & Dolan 1996; Goldenberg, Mullbacher & Nowak 1995; Howard, ffytche, Barnes, McKeefry, Ha, Woodruff, Bullmore, Simmons, Williams, David, & Brammer 1998; Mellet, Petit, Mazoyer, Denis, & Tzourio 1998; Mellet, Tzourio, Crivello, Joliot, Denis, & Mazoyer 1996; Roland & Gulyas 1994b; Roland & Gulyas 1995; Silbersweig & Stern 1998); but see the review in (Farah 1995) and the some of the published debate on this topic (Farah 1994; Roland & Gulyas 1994a, 1994b). Other evidence comes from clinical cases of brain damage and is even less univocal in supporting the involvement in mental imagery of the earliest, topographically organized areas of visual cortex (Roland & Gulyas 1994b). There is some reason to think that the activity associated with mental imagery occurs at many loci, including higher-levels of the visual stream (Mellet et al. 1998).

Despite the weight placed on neural imaging studies by proponents of the picture theory, the involvement of visual areas of cortex – even of topographically organized areas of early vision – in mental imagery would not itself support a cortical display view of imagery. In order to support such a view it is important not only that such topographically organized areas be involved in imagery, but also that their involvement be of the right sort – that the way their topographical organization is involved reflects the spatial properties of the image, particularly as the latter is experienced and as it is assumed to function in accounting for the many imagery findings in the literature. Very few neuroscience studies
meet this criterion, even when they show that early visual areas are activated during mental imagery. That is because such studies generally do not provide any evidence concerning how spatial information is mapped in the visual cortex. One of the few examples of a study that does address this question was reported in (Kosslyn et al. 1995). Unlike most neural imaging studies this one relates a specific spatial property of a phenomenal mental image (its size) to a pattern of neural activity. Thus it behooves us to look at the findings in detail.

The (Kosslyn et al. 1995) study showed that “smaller” mental images (mental images that the observer subjectively experiences as occupying a smaller portion of the available “mental display”) are associated with more activity in the posterior part of the medial occipital region while “larger” images are associated with more activity in the anterior parts of the region. Since this pattern is similar to the pattern of activation produced by small and large retinal images, respectively, this finding was taken to support the claim that not only does activation of visual cortical areas occur during mental imagery, but also that the form of the activation is the same as that which derives from vision. This, in turn is interpreted as showing that imagery creates a cortical display which maps represented space onto cortical space. Because of this (Kosslyn et al. 1995, p. 496) feel entitled to conclude that the findings “indicate that visual mental imagery involves ‘depictive’ representations, not solely language-like descriptions.”

But notice that even if the cortical activity monitored by PET scans corresponded to a mental image, the evidence only showed that a larger mental image involves activity that is located where parts of larger retinal images tend to project – that is at best they show that larger mental images involve activity in different locations in the cortex than the activity associated with smaller mental images. We have good reason to believe that the reason larger retinal images activate the regions they do (in the case of vision) is because of the way that the visual pathway projects from the periphery of the retina to the occipital cortex (Fox, Mintun, Raichle, Miezin, Allman, & Van Essen 1986). This pattern of activation does not map the size of the image onto a metrical spatial property of its cortical representation. In particular, it is important that the data do not show that image size is mapped onto the size of the active cortical region, as would be required by the cortical display view, and as would be required if images had the property that they “preserve metrical information” as claimed by Kosslyn and others. It is also not the pattern that is required in order to account for the imagery data reviewed earlier. For example, the explanation for why it takes less time to notice features in a large image is supposed to be that it is easier to discern details in a large cortical display, not that the image is located in the more anterior part of the medial occipital cortex. The property of being located in one part of the visual cortex rather than another simply does not bear on any of the behavioral evidence regarding the spatial nature of mental images discussed earlier. Consequently the finding cannot be interpreted as supporting the cortical display view of mental imagery, nor does it in any way help to make the case for a literal-space view or for the picture theory.

Those of us who eschew dualism naturally assume that something different happens in the brain when a different phenomenal experience occurs. Consequently, something different must occur in the brain when a larger image is experienced (this is called the supervenience assumption and few scientists would dispute it). The point has never been to question materialism, but only to question the claim that the content of an image maps onto the brain in a way that helps explain the imagery results (e.g., mental scanning times, image-size effects, etc) and perhaps even the subjective content of mental images. Discovering that the a larger phenomenal image mapped onto a larger region of brain activity might have provided some support for this view, since it might at least have suggested a possible account for such findings as that it takes longer to scan larger image distances and that it takes longer to see details in smaller images. But finding that a larger mental image merely activated a different area of the brain is no help in this regard. Incidentally, while the data tend to confirm that the pattern of cortical activity changes in similar ways for similar differences in perceived and imagined patterns, this does not support the existence of a cortical display in either imagery or vision. And that’s as we would expect, given that there is no evidence that an extended display is involved in visual processing beyond the retina and its
primary projection (but see Note 14). The hope of enlisting neuroscience to provide support for a picture theory of mental imagery by showing an overlap between vision and imagery rests, in the first instance, on the acceptance of a false theory of vision.

A somewhat different kind of evidence for the neural basis of image size was reported by (Farah, Soso & Dasheiff 1992) based on a clinical case. Farah et al. reported that a patient who developed tunnel vision after unilateral occipital lobectomy also developed tunnel imagery. Once again this may well show that the two cases of tunnel vision have a common underlying neural basis, and that this basis may be connected with the visual system. But it does not show that this basis has anything to do with a topographical mapping of the spatial property of images onto spatial properties of a neural display. I might also point out that other possible explanations for the Farah et al. finding are possible which do not assume that the surgery resulted in damage to the imagery system. If, as suggested earlier, many of the properties of mental images actually arise from the implicit task requirement of simulating what things would look like in the corresponding perceptual situation, then a plausible explanation for the Farah et al. finding is that the patient was simply demonstrating a tacit knowledge of what the post-surgery world looked like to her. In the (Farah et al. 1992) study the patient had nearly a year of post-surgery recovery time before the imagery testing took place. During this time she would certainly have become familiar with what things looked like to her, and would therefore have been in a position to simulate her visual experience by demonstrating the relevant phenomena when asked to image certain things (e.g., to answer appropriately when asked at what distance an image of a horse would overflow her image). This would not be a case of the patient being disingenuous or being influenced by what the experimenters expected to find, which Farah et al. were at pains to deny, but of doing her best to carry out the task required of her, namely to “imagine how it would look.”

7.2 What would it mean if all the neuroscience claims turned out to be true?

Results such as those of Kosslyn et al. and Farah et al. have been widely interpreted as showing that retinotopic picture-like displays are generated on the surface of the visual cortex during imagery and that it is by means of this spatial display that images are processed, patterns perceived from mental images and the results of mental imagery experiments produced. In other words these results have been taken to support the view that mental images are literally two-dimensional displays projected onto primary visual cortex. I have already suggested some reasons why the neuroscience evidence does not warrant such a strong conclusion (and that a weaker “functional space” assumption is incapable of explaining the results of mental imagery studies without extrinsic constraints that are compatible with, and could be applied to, any form of representation). In addition, it should be remembered that standing against this interpretation of the neuroscience findings, is a large body of behavioral evidence that cannot be ignored. If we are to take seriously the conclusions suggested by the researchers who use neuroscience evidence to argue for the picture-theory (or the cortical-display theory), we need to understand the role that could be played by a literal picture being projected onto visual cortex.

Here is a summary of some reasons to doubt the assumption that a picture is projected onto the visual cortex when we entertain mental images.

1. There is a great deal of evidence that the capacity for visual imagery is independent of the capacity for visual perception. It is hard to reconcile the view that it is the topographical form of activation in early visual areas that is responsible for the visual image results discussed earlier, given the evidence for the dissociation between imagery and visual perception. For example, there is convergent evidence for the dissociation between the capacity for visual imagery and such visual deficits as cortical blindness (Chatterjee & Southwood 1995; Dalman, Verhagen & Huygen 1997; Goldenberg et al. 1995; Shuren, Brott, Scheff, & Houston 1996), dyschromatopsia (Bartolomeo, Bachoud-levi & Denes 1997; De Vreese 1991; Howard et al. 1998), visual agnosia (Behrmann, Moscovitch & Winocur 1994; Behrmann, Winocur & Moscovitch 1992; Jankowiak, Kinsbourne, Shalev, & Bachman 1992; Servos & Goodale 1995) and visual neglect (Beschin, Cocchini, Della Sala, & Logie 1997). The independence of imagery
and vision is also supported by a wide range of both brain-damage and neuroimaging data and the dissociation has been shown in both directions. The case for independence is made all the stronger by the evidence that blind people show virtually all the skills and psychophysical phenomena associated with mental imagery (including the reaction-time data discussed in section 3) (Barolo, Masini & Antonietti 1990; Carpenter & Eisenberg 1978; Cornoldi, Bertuccelli, Rocchi, & Sbrana 1993; Cornoldi, Calore & Pra-Baldu 1979; Craig 1973; Dauterman 1973; Dodds 1983; Easton & Bentzen 1987; Hampson & Duffy 1984; Hans 1974; Heller & Kennedy 1990; Johnson 1980; Jonides, Kahn & Rozin 1975; Kerr 1983; Marmor & Zaback 1976; Zimler & Keenan 1983). Blind people may even report a comparable phenomenology concerning object shape as that of sighted people. While there have been attempts to explain these dissociations by attributing some of the lack of overlap to an “image generation” phase that is presumably involved only in imagery (see the recent review in, Behrmann 2000), this image-generation proposal does not account for much of the evidence for the independence of imagery and vision, in particular, it cannot explain how one can have spared imagery in the presence of such visual impairments as total cortical blindness. It has also been suggested that what characterizes patients who show a deficit on certain imagery-generation tasks (e.g., imagining the color of an object) is that they lack the relevant knowledge of the appearance of objects (Goldenberg 1992; Goldenberg & Artner 1991). Consequently, insofar as blind people know (in a factual way) what objects are like (including aspects of their “appearance” – such as their size, shape and orientation) it is not surprising that they should exhibit some of the same psychophysical behaviors in relation to these properties.

Of course it is also possible that many cortically blind people have deficits in only some parts of their visual system, and in particular, that they have damage to the more peripheral parts of the visual system, parts that are closer to sensory neurons. Thus it might be that they still have the use of other parts of their visual system where the input is from a more central cortical locus. While this is certainly a possibility, it is not compatible with the view that in both vision and imagery images are projected onto the primary visual cortex, since cortical blindness invariably involves damage to the visual cortex. A more plausible alternative may be that imagery studies do not tap a specifically visual capacity but a more general spatial capacity.

There is good reason to believe that blind subjects have normal spatial abilities – and indeed blind children show the same spontaneous acquisition and use of a spatial vocabulary as do sighted children (Landau & Gleitman 1985). Thus mental imagery might involve a spatial mechanism rather than a visual one. This possibility, however, provides no comfort to the picture-theorists and it also leaves open the question of why these mechanisms connect with the motor system in a different manner than when they are visually stimulated (see section 5.4).

2. The conclusion that many people have drawn from the neural imaging evidence cited earlier, as well as from the retinotopic nature of the areas that are activated, is that images are two-dimensional retinotopic displays. But that can’t be literally the case. If mental images are depictive they would have to be three-dimensional inasmuch as the phenomenology is that of seeing a three-dimensional scene. Moreover, similar mental scanning results are obtained in depth as in 2D (Pinker 1980) and the phenomenon of “mental rotation” is indifferent as to whether rotation occurs in the plane of the display or in depth (Shepard & Metzler 1971). Neither can the retinotopic “display” in visual cortex literally be three-dimensional. The spatial properties of the perceived world are not reflected in a volumetric topographical organization in the brain: as one penetrates deeper into the columnar structure of the cortical surface one does not find a representation of the third dimension of the scene. Furthermore, images represent other properties besides the spatial ones. For example, they represent color and luminance and motion. Are these also to be found displayed on the surface of the visual cortex? If not, how do we reconcile the apparently direct spatial mapping of 2D spatial properties with a completely different form of mapping for depth and for other contents of images of which we are equally vividly aware?

3. The cortical display view of mental imagery assumes that mental images consist not only in the
activation of a pattern that is the same as the pattern activated by the corresponding visual percept, but it also assumes that such a pattern mimics the retinotopic projection of a corresponding visual scene. In some versions it even assumes that the pattern displayed in the cortex is a spatial extension of such a retinal projection, which incorporates a larger region of the scene than that covered by the retina, thereby explaining the subjective impression we have of a stable panoramic view of the world as we move our eyes around. Among the arguments put forward for the existence of an inner display is that it is needed to explain stability of the percept over eye movements and the invariance of recognition with translations of the retinal image (Kosslyn 1994, Chapter 4). It has also been suggested that the display is needed to account for the completion of the apparently detailed visual percept in the face of highly incomplete and partial sensory data (Kosslyn & Sussman 1995). The assumption behind these arguments is that incomplete sensory data are augmented in a visual display before being given over to the visual system responsible for recognition (and which presumably leads to our conscious awareness). While it may be that neural processes are responsible for certain cases of “filling in” phenomena, it is also clear that they do not do so by completing a partially filled display (for a sophisticated discussion of the issue of filling-in, which makes it clear that cases of neural completion do not imply “analytical isomorphism,” see Pessoa, Thompson & Noë 1998).

Notwithstanding the fact that the early part of the visual cortex appears to be organized retinotopically, it is highly unlikely that this retinotopic organization serves to shield the inner eye from the incompleteness and instability of the incoming sensory data, and thereby gives rise to such properties as the invariance of recognition over different retinal locations. There is every reason to believe that vision does not achieve stability and completeness, despite rapidly changing and highly partial information from the sensors, by accumulating the information in a spatially extended internal display. In fact that we sometimes feel we are examining an internal display in vision is simply a mistaken inference. Even if we had such a display we would not see it, we see the world and it is the world we see that appears to us in a certain way. The evidence clearly shows that the assumption that visual stability and saccadic integration is mediated by an inner-display is untenable (Blackmore et al. 1995; Irwin 1991; McConkie & Currie 1996; O'Regan 1992) since information from successive fixations cannot be superimposed in a central image as required by this view. Recent work on change blindness also shows that the visual system encodes surprisingly little information about a scene between fixations, unless attention has been drawn to it (Rensink 2000a; Rensink et al. 1997; Simons 1996), so there is no detailed pictorial display of any kind in vision, let alone a panoramic one.

4. Although we can reach for imagined objects there are significant differences between the way our motor system interacts with vision compared with the way it interacts with mental imagery (Goodale et al. 1994), as we saw in section 5.4. Such differences provide strong reasons to doubt that imagery provides input into the dorsal stream of the early vision system where the visuomotor control process begins, as it would if it were a retinotopic cortical projection.

5. Accessing information from a mental image is very different from accessing information from a scene. To take just one simple example, we can move our gaze as well as make covert attention movements relatively freely about a scene, but not on a mental image. Try writing down a 3 x 3 matrix of random letters and read them in various orders. Now imagine the matrix and try doing the same with it. Or, for that matter, try spelling a familiar word backwards by imagining it written. Unlike the 2D matrix, some orders (e.g., the diagonal from the bottom left to the top right cell) are extremely difficult to scan on the image. If one scans one’s image the way it is alleged one does in the mental scanning experiments, there is no reason why one should not be able to scan the matrix freely. Of course one can always account for these phenomena by positing various properties specific to images generated by cognitive processes, as opposed to ones we retain from short-term visual memory. For example, one might assume that there is a limit on the number of elements that can be generated at one time, or one might assume that elements decay. But such assumptions are completely ad hoc. Visual information does not appear to fade as fast and in the same way from images held in short-term visual memory (Ishai & Sagi 1995), nor does it appear to fade in the case of images used to investigate mental scanning
phenomena (which are much more complex, as shown in Figure 1). Moreover the hypothesized fading rates of different parts of an image have to be tuned post hoc to account for the fact that it is the conceptual as opposed to the graphical complexity of the image determines how the image can be read and manipulated (that is, to account for the fact that what one sees the image as, how one interprets it, rather than its geometry, is what determines its apparent fading). For example, it is the conceptual complexity of images that matters in determining the difficult of an image superposition task (Palmer 1977), or in determining how quickly figures are “mentally rotated” (Pylyshyn 1979).

6. The central role of conceptual, as opposed to graphical properties of an image, alluded to above, is an extremely general and important property of images. It relates to the question of how images are distorted or transformed over time, to how mental images can or can’t be (re)interpreted, and to how they can fail to be determinate in ways that no picture can fail to be determinate (Pylyshyn 1973, 1978, 1984). For example, no picture can fail to have a size or shape or can fail to indicate which of two adjacent items is to the left and which to the right, or can fail to have exactly \( n \) objects (for some \( n \)), whereas mental images can be indeterminate in many ways. Not surprisingly, there are many ways of patching up a picture theory to accommodate such findings. For example one can add assumptions about how images are tagged as having certain properties (perhaps including the property of not being based on real perception) and how they have to be incrementally refreshed from non-image information stored in memory, and so on, thus providing a way to bring in conceptual complexity and indeterminacy through the image generation function. With each of these accommodations, however, one gives the actual image less and less of an explanatory role until eventually one reaches the point where the display becomes a mere shadow of the mechanism that does its work elsewhere, as when the behavior of an animated computer display is determined by an extrinsic encoding of the principles that govern the animation rather than by intrinsic properties of the display itself.

7. The visual appearance of information projected onto a retinotopic display is very different from the appearance of information in a mental image. Images on the retina, and presumably on the retinotopically-mapped visual cortex, are subject to Emmert’s law: Retinotopic images superimposed onto a visual scene change their apparent size depending on the distance of the background against which they are viewed. Mental images imagined over a perceived scene do not change their size as the background recedes, providing strong evidence that they are not actually projected onto the retinotopic layers of the cortex.

8. Images do not have the signature properties of early vision (such as the properties discussed in Hochberg 1968). If we create mental images from descriptions we do not find such phenomena as spontaneous interpretation of certain 2D shapes as representing 3D objects, spontaneous reversals of bistable figures, amodal completion or subjective contours (Slezak 1995), visual illusions, as well as the incremental construction of visual interpretations and reinterpretations over time, as different aspects are noticed. There is even evidence (discussed in section 6.4) that such early vision phenomena as motion aftereffects do not affect imagined motion the same way that they affect real perceived motion.

7.3 Is the ‘mind’s eye’ just like a real eye?

Here is another way to think about the question of whether mental images could plausibly consist of patterns projected onto the cortex. Suppose it turns out that that when we entertain a mental image there is an actual copy of that very image (say in the form of neural activity) on the surface of primary visual cortex (or, for that matter, on the retina itself; the conceptual issue would be the same in either case). What would that tell us about the nature and role of mental images in cognition? We have known at least since Descartes that there is an image on our retina in visual perception, and perhaps there is also some transformed version of this image on our cortex, yet knowing this has not made us any wiser about how visual perception works. Indeed, ruminating on the existence of an image has only raised such problems as why we do not see the world as upside down, given that the image on the retina is upside down. The temptation to assume a literal picture observed through a “mind’s eye” may be very strong
but it leads us at every turn into blind alleys.

For example, some of the psychophysical evidence that is cited in support of a picture theory of mental imagery suggests a similarity between the mind’s eye and the real eye that is so remarkable that it ought to be an embarrassment to picture-theories. It not only suggests that the visual system is involved in imagery and that it examines a pictorial display, but it appears to attribute to the “mind’s eye” many of the properties of our own eyes. For example, it seems that the mind’s eye has a visual angle like that of a real eye (Kosslyn 1978) and that it has a field of resolution which is roughly the same as our eyes; it drops off with eccentricity according to the same function and inscribes a similar elliptical acuity profile as that of our eye (Finke & Kosslyn 1980; Finke & Kurtzman 1981a). It even appears that the “mind’s eye” exhibits the “oblique effect” in which the discriminability of closely-spaced horizontal and vertical lines is superior to that of oblique lines (Kosslyn, Sukel & Bly 1999b). Since in the case of the eye, such properties arise from the structure of our retina and of its projection onto the visual cortex, it would appear to suggest that the mind’s eye is similarly constructed. Does the mind’s eye then have the same color profile as that of our eyes – and perhaps a blind spot as well? Does it exhibit after-images? And would you be surprised if experiments showed that it did? Of course, the observed parallels could be just coincidence, or it could be that the distribution of neurons and connections in the visual cortex has come to reflect the type of information it receives from the eye. But it is also possible that such phenomena reflect what people have implicitly come to know about how things appear to them, a knowledge which the experiments invite them to use in simulating what would happen in a visual situation that parallels the imagined one. Such a possibility is made all the more plausible in view of the fact that the instructions in these imagery experiments explicitly ask observers to “imagine” a certain visual situation – that is, to imagine that they are in a certain visual circumstances and to imagine what it would look like to see things, say, in their peripheral vision. (I have often wondered whether people who wear thick-framed glasses would have a smaller field of vision in their mind’s eye).

The picture that we are being presented, of a mind’s eye gazing upon a display projected onto the visual cortex, is one that should arouse our suspicion. It comes uncomfortably close to the idea that properties of the external world, as well as of the process of vision (including the resolution pattern of the retina and the necessity of moving one’s eyes around the display to foveate features of interest), are internalized in the imagery system. If such properties were built into the architecture, our imagery would not be as plastic and cognitively penetrable as it is. If the “mind’s eye” really had to move around in its socket we would not be able to jump from place to place in extracting information from our image the way we can. And if images really were pictures on the cortex, the necessity of a homunculus to interpret them would not have been discharged, notwithstanding claims that such a system had been implemented on a computer. Computer implementation does not guarantee that what is said about the system, viewed as a model of the mind/brain, is true. Nor does it guarantee that the theory it implements is free of the assumption that there is an intelligent agent in one of the boxes. As (Slezak 1995) has pointed out, labels on boxes in a computational model are not merely mnemonic; the choice of a label often constitutes a substantive claim that must be independently justified. Labeling a box as, say “attention,” (as is done in the model described in Kosslyn et al. 1979) may well introduce a homunculus into the theory, despite the fact that the system is implemented as a running program which generates some of the correct predictions in a very limited domain. That’s because the label implies that the performance of the system will continue to mirror human performance in a much broader domain; it implies that the system can be scaled up in ways that are consistent with the assigned label.

7.4 What has recent neuroscience evidence done for the “imagery debate”?

Where, then, does the “imagery debate” stand at present? In the first place, although many investigators (including Kosslyn 1994, Chapter 1) write as though recent neuroscience evidence supercedes all previous behavioral evidence, nothing could be further from the truth. It was behavior (and phenomenological) considerations that raised the puzzle about mental imagery in the first place and that suggested the picture theory. And it is a careful consideration of that evidence and its alternative
interpretations that have cast doubt on the picture theory. Consequently, even if real colored stereo pictures were found on the visual cortex, the problems raised thus far in this article would remain and would continue to stand as evidence that such cortical pictures were not serving the function attributed to them. For example, the fact that phenomena such as mental scanning are cognitively penetrable is strong evidence that whatever is displayed on the cortex is not what is responsible for the patterns of behavior observed in mental imagery studies. The mere fact that the data are biological does not give it a privileged status in deciding the truth of a psychological theory, especially one whose conceptual foundations are already shaky.

As I suggested near the beginning of this article, where the “imagery debate” stands today depends on what you think the debate was about. If it was supposed to be about whether reasoning using mental imagery is somehow different from reasoning without it, nobody can doubt it. If it was about whether in some sense imagery involves the visual system, the answer there too must be affirmative, since imagery involves similar experiences to those produced by (and, as far as we know, only by) activity in some part of the visual system (though not in V1, according to Crick & Koch 1995). The real question is; in what way is the visual system involved and what does that tell us about the properties of mental imagery, and about how the mind generates and uses images? It is much too early and much too simplistic to claim that the way the vision system is deployed in visual imagery is by allowing us to look at a reconstructed retinotopic input of the sort that comes from the eye (or at least to some locally-affine mapping of this input).

Is the debate, as (Kosslyn 1994) claims, about whether images are depictive as opposed to descriptive? That all depends on what you mean by “depictive.” Is any representation of geometrical, spatial, metrical or visual properties depictive? If that makes it depictive then any description of how something looks is thereby depictive. Does being depictive require that the representation be organized spatially? As I suggested, that depends on what restrictions are placed on “being organized spatially”; most forms of representation, including symbol structures, use different spatial locations to distinguish among represented individuals. Does being depictive require that images “preserve metrical spatial information” as has been claimed (Kosslyn et al. 1978)? Again that depends on what it means to “preserve” metrical space. If it means that the image must represent metrical spatial information, then any form of representation will have to do that to the extent that it can be shown that people do encode and recall such information. But any system of numerals, as well any analogue medium, can represent magnitudes in a useful way. If the claim that images preserve metrical spatial information means that imagery uses spatial magnitudes to represent spatial magnitudes, then this is a form of the literal picture theory, which I have argued is not supported by the evidence.

The neuroscience evidence we briefly looked at, while interesting in its own right, does not appear capable of resolving the issue about the nature of mental images, largely because the questions have not been formulated appropriately and the options are not well understood – with the single exception of the literal cortical display theory which turns out to be empirically inadequate in many different ways. One major problem with providing a satisfactory theory of mental imagery is that we are attempting to account not only for certain behavioral and neuroscience findings, but we are attempting to do so in a way that remains faithful to certain intuitions and subjective experiences. It is not obvious that all these constraints can be satisfied simultaneously. There is no a priori reason why an adequate theory of mental imagery will map onto conscious experience in any direct and satisfactory way. Indeed if the experience in other sciences and in other parts of cognitive science is any indication, the eventual theory will not do justice to the content of our subjective experience and we will simply have to live with that fact the way physics has had to live with the fact that the mystery of action-at-a-distance does not have a reductive explanation.

7.5 Is the “picture theorist” a straw man?

The typical response to arguments such as those raised in this section is that it takes the picture
theory too literally and nobody really believes that there is an actual 2D display in the brain. For example, (Denis & Kosslyn 1999) maintain that “No claim was made that visual images themselves have spatial extent, or that they occupy metrically defined portions of the brain” and (Kosslyn 1994, p329) admits that “images contain ‘previously digested’ information.” But if that is the case, how does one explain the increase in time to scan greater image distances or to report details in smaller images? An explanation of these phenomena that appeals to a depictive representation requires a literal sense of ‘spatial extent ’ otherwise the explanation does not distinguish the depictive story from what I have called the null hypothesis (see the discussion of the ‘functional space’ alternative in section 5.2). If one denies the literal view of a cortical display, how does one interpret the claim that activation of topographically organized areas of the visual cortex during imagery establishes that images are “depictive”? If one were looking in the brain for evidence of a “functional space”, what exactly would one look for? It is because picture theorists are searching for a literal 2D display that the research has focused on showing imagery-related activity in cortical Area 17.

The view that is favored by picture theorists is clearly illustrated by the importance that has been attached to the finding described in (Tootell, Silverman, Switkes, & de Valois 1982). In this study, macaques were trained to stare at the center of a pattern of flashing lights, while the monkeys were injected with radioactively tagged 2-deoxydextroglucose (2-DG), whose absorption is related to metabolic activity. Then the doomed animal was sacrificed and a record of 2-DG absorption in its cortex was developed. This record showed a retinotopic pattern in V1, which corresponded closely to the pattern of lights. In other words, it showed a picture in visual cortex of the pattern that the monkey had received on its retina, written in the ink of metabolic activity. This led some people to conclude that we now know that a picture in primary visual cortex appears during visual perception and is the basis for visual perception [14]. Although no such maps have been found for imagery, there can be no doubt that this is what the picture-theorists believe is there and is responsible for both the imagery experience and the empirical findings reported when mental images are being used. People who have accepted this line of argument are well represented in the imagery debate: They are not “straw men”!

The problem is that while the literal picture-theory or cortical display theory is what provides the explanatory force and the intuitive appeal, it is always the picture metaphor that people retreat to in the face of the implausibility of the literal version of the picture theory. This is the strategy of claiming a decisive advantage for the depictive theory because it has the properties cited in the quotation in section 5.1 (e.g., it resembles what it represents), it is located in the topographically organized areas of visual cortex, it “preserves metrical information” and so on; then, in the face of its implausibility, systematically retreating away from the part of the claim that is doing the work – the literal spatial layout.

8 Conclusion: What is special about mental imagery?

The theme that has run through this essay is that we have thus far not been given adequate reasons to reject the null hypothesis and to accept that what goes on in mental imagery is in any way like examining a picture. [15] Yet the conclusion that reasoning using imagery is the same as reasoning that is not accompanied by the experience of “seeing in the mind’s eye” is surely premature. It suggests that the fact that we have certain phenomenal experiences is irrelevant to understanding the nature of images, or that the image experience is “epiphenomenal”, [16] neither of which is warranted, even if at the present time we have no adequate understanding of what role the conscious experience of image content might play (see section 6.1). So what exactly are we entitled to conclude? What I have argued here is primarily that we are not entitled to take the tempting road of assuming that the content of our experience reflects in any direct way the nature of our cognitive information processing activity (in other words, one ought to guard scrupulously against what Pessoa et al. 1998, call the “analytical isomorphism” assumption). Thus one ought to start off with strong skepticism about the idea that
images are like pictures.

The search for a system of representation that retains some of the attractive features of pictures and yet can serve as the basis for reasoning has been the holy grail of many research programs, both in cognitive science and in artificial intelligence. The hope has been that one might develop a coherent proposal that captures some of what is special about reasoning with images, without succumbing to the Cartesian Theater trap. One way to approach this problem is to consider the most general constraints or boundary conditions that have to be met by a system of imagistic reasoning. Even if we do not have a detailed theory of the form of representation underlying imagistic thoughts, we do know some of the conditions it must meet. Starting by setting conditions on an adequate theory is not a new strategy. It was extremely successful in linguistics (Chomsky 1957a; Chomsky 1957b) and is routine in theoretical physics (“thought experiments” can be viewed as explorations of what is entailed by such general conditions). (Newell & Simon 1976) have also remarked on the importance of such orienting perspectives (which they called “laws of qualitative structure”) for scientific progress.

As an example of such constraints, Fodor and I have argued (Fodor & Pylyshyn 1988) that in order to be adequate as vehicles of reasoning, such representations must meet the conditions of productivity, compositionality, and systematicity. Representations underlying imagery should meet additional conditions as well. Some years ago I proposed several possible constraints that are specific to mental images (Pylyshyn 1978). These have nothing to do with how images appear. Rather they focus on the idea that mental images represent potentially visible token individuals or small sets of individuals. Because they represent individuals, they do not explicitly encode such set-properties as the cardinality of the set of individuals (they do not explicitly encode facts such as that there are eight boxes, or universally quantified propositions such as “all Xs are Y”). Because they represent individuals, they in effect assert the presence of some individuals or properties, and not their absence (e.g., they cannot represent a prepositional content such as “there is no X” or “it is not the case that P”). In addition, the content of images tends to involve visual rather than abstract properties.

The only theory I am aware of that shares some of the formal properties listed above is a system of formal reasoning developed by (Levesque & Brachman 1985). Levesque discovered a fundamental trade-off between the expressive power of a system of representation and the complexity of drawing inferences in that system. In (Levesque 1986) he describes an expressively weaker form of logical representation (which he calls a “vivid Representation”) that allows inferences to be drawn essentially by pattern matching. As in my earlier speculation about what is special about mental imagery, representations in this system do not permit the direct expression of negation (e.g., the only way that they can represent the proposition “there are no red squares” is by representing a scene that contains no red squares), or disjunction (e.g., they can only represent the proposition “the squares are either red or large” by allowing two possible representations, one with red squares and one with large squares), and they do not allow universal quantification (e.g. they can only represent the proposition “all squares are red” by explicitly representing each square, however many there are, and showing each as red). A vivid representation can only express the fact that there are 5 objects in a scene by representing each of the objects (of which there would be 5 in all). Levesque then proves some remarkable complexity properties for databases consisting of such vivid representations. Even though this work was not directly motivated by the imagery debate, it has the virtue of meeting the boundary conditions on an adequate system of representation; it does not postulate properties that we know are inadequate for the representation of knowledge, such as literally spatial displays.
References


Psychologica, 67-79.

Notes

∗ I wish to thank Jerry Fodor, Ned Block and Peter Slezak for useful exchanges, and reviewers Michael McClosky, David Marks, Art Glenberg, and Mel Goodale, for providing helpful comments on an earlier draft. This work was supported by NIH Grant 1R01-MH60924.

[1] Some of these demonstrations can be viewed by downloading Quicktime© animations from: http://www.cbr.com/~rensink/flicker/flickDescr.html

[2] When we first carried out these studies we were criticized (quite rightly, in my view) on the grounds that it was obvious that you did not have to scan your image if you did not want to, and if you did you could do so according to whatever temporal pattern you chose. It still seems to me that the studies we carried out only demonstrate the obvious. That being the case one might wonder what the great fuss (and is) was about over the scanning phenomenon (as well as the image size phenomenon describe below); why dozens of studies have been done on it and why it is interpreted as showing anything about the nature of mind as opposed to choices that subjects make.

[3] People have suggested that one can accommodate this result by noting that the observed phenomenon depends on both the form of the image and on the particular processes that use it, so that the differences in the process can account for the different result obtained with different tasks (e.g., in this case attention might be moved by a
“jump” operation). In that case the assumption that there is a depictive representation that “preserves metrical distance” does not play a role. The problem then becomes to specify the conditions under which the spatial character of the image does or does not play a role. A plausible answer is that scanning results are obtained when the subject thinks that imagining scanning a display is part of the task at hand. Of course these are also the conditions under which the subject understands the task to be the simulation of visual scanning and thus recreating the time-distance scanning effect.

[4] I don’t mean to pick on Stephen Kosslyn, who (along with Allan Paivio and Roger Shepard) has done a great deal to promote the scientific study of mental imagery. I focus on Kosslyn’s work here because he has provided what is currently the most highly developed and explicit theory of mental imagery and has tried to be particularly explicit about his assumptions, and also because his work has been extremely influential in shaping psychologists’ views about the nature of mental imagery. In that respect his views can be taken as the received view in much of the field.

[5] If we look closely at what goes on in a computer implementation of a matrix, we see even more clearly that some of the space-like features are only in the mind of the user. For example, the matrix is said to contain a representation of empty space, in that the cells between features are actually represented explicitly. Whether registers are actually reserved in a computer for such empty cells is a matter of implementation and need not always be the case – indeed it often is not the case in efficient implementations of sparse matrices. Moreover, since an empty place is just a variable with no value (or a default zero value), any form of representation can make the assumption that there are names for unfilled places. In fact you don’t even have to assume that such place names exist prior to an inquiry being made about their contents; names can be (and are) created on the fly as needed (e.g., using LISP’s Gensym function). The same goes for the apparent pairs of numbers we think of as matrix coordinates; these are mapped onto individual names before being used to retrieve cell contents. The point is not just that the implementation betrays the assumption that such properties are inherent, it is also that how a matrix functions can be just as naturally viewed non-spatially since a matrix is not required by any computational constraints to have the properties assumed in a table display.

[6] There is one other way of interpreting a functional space such as associated with a matrix. Rather than viewing it as a model of (real, physical) space it might be thought of as a model of a (real, physical) analogue system that itself is an approximate analogue of space. In order for there to be an analogue model of space in the brain, however, there would have to be a system of brain properties that instantiate at least a local approximation of the Euclidean axioms. It would not do to just have an analogue representation of some metrical properties such as distance, which itself is an eminently reasonable assumption, but we would need a whole system of such physically instantiated analogue properties in the brain. As in any analogue model, there would have to be a
well-defined homomorphism from a set of spatial properties to a set of analogue properties: it would have to be possible to define predicates like between, adjacent, collinear, and so on as well as the operation of moving-through these analogue dimensions at a specified speed. As far as I know, nobody has seriously developed a proposal for such an analogue representation (although the work of Nicod 1970, might be viewed as a step towards such a goal). But from the perspective of the present thesis this alternative suffers from the same deficiency that a literal spatial proposal does: It fails to account for the cognitive penetrability of the empirical phenomena that are cited in support of the picture theory of mental imagery.

[7] There are almost always some visual elements, such as “textons,” (Julesz 1981), that can serve as indexed objects. It is dubious whether places unoccupied by any visible feature can be indexed (which is why vision in a featureless environment is so unstable, Avant 1965). The one possibility, suggested by the work of (Taylor 1961), is that locations clearly definable in terms of nearby visible objects can be indexed. Taylor showed that observers can encode locations more accurately when they are easily specified in relation to some visible anchors (such as a being the “midpoint” between two visible objects).

[8] (Brandt & Stark 1997) reported that the sequence of eye movements observed when inspecting a mental image inscribe similar scan paths to those observed when inspecting an actual display. But since the experiment was not carried out in total darkness, the eye movements could have been made to the faint visible cues, rather than to image features (see section 5.3). In addition, since moving one’s eye does not result in viewing different parts of an inner picture, the contents of the display could not provide the feedback required to control a sequence of eye movement beyond the initial ballistic saccade, so that in any case the sequence of eye movements during imagery are likely due to a different mechanism than the one that controls the sequence of eye movements in vision.

[9] Finke also found a significant effect of “vividness” of imagery, as determined from the Vividness of Visual Imagery Questionnaire (VVIQ) (Marks 1973), with the adaptation effect being much higher for vivid imagers. It is not clear how to interpret such a result, however, given that subjects who were high in vividness also had significantly higher adaptation scores in the control condition where there was no feedback (visual or imaginal) about errors of movement. These findings (along with the connection between vividness and hypnotic suggestibility reported by Crawford 1996; Glisky, Tataryn & Kihlstrom 1995; Kunzendorf, Spanos & Wallace 1996) increase the likelihood that experimental demand effects may be involved in the performance patterns of high-scoring subjects.

[10] The conditions under which one gets more or less adaptation is discussed in (Howard 1982). The most important requirement is that the discordant information be
salient for the subject and that it be interpreted as a discordance between two measures of the position of the same limb. Thus anything that focuses more attention on the discordance and produces greater conviction that something is awry helps strengthen the adaptation effect. Thus it is not surprising that merely telling subjects where their hand is does not produce the same degree of adaptation as asking them to pretend that it actually is at a particular location, which is what imagery instructions do.

[11] Many of these studies have serious methodological problems, which we will not discuss here in detail. For example, a number of investigators have raised questions about many of these illusions (Predebon & Wenderoth 1985; Reisberg & Morris 1985) where the likelihood of experimenter demand is high. The usual precautions against experimenter influence on this highly subjective measure were not taken (e.g., the experiments were not done using a double-blind procedure). The most remarkable of the illusions, the orientation-contingent color aftereffect, known as the McCollough effect, is perhaps less likely to lead to an experimenter-demand effect since not many people know of the phenomenon. Yet (Finke & Schmidt 1977) reported that this effect is obtained when part of the input (a grid of lines) is merely imagined over the top of a visible colored background. But the Finke finding has been subject to a variety of interpretations as well as to criticisms on methodological grounds (Broerse & Crassini 1981, 1984; Harris 1982; Kunen & May 1980, 1981; Zhou & May 1993) so will not be reviewed here. Finke himself (Finke 1989) appears to accept that the mechanism for the effect may be that of classical conditioning rather than a specifically visual mechanism.

[12] In a recent paper (Kosslyn et al. 1999a) also claimed that if area 17 is temporarily impaired using repetitive transcranial magnetic stimulation (rTMS), performance on an imagery task is adversely affected (relative to the condition when they do not receive rTMS), suggesting that the activation of area 17 may be not only correlational, but may also play a causal role. However, this result must be treated as highly provisional since the nature and scope of the disruption produced by the rTMS is not well established and the study in question lacks the appropriate controls for this critical question; in particular, there is no control condition measuring the decrement in performance for comparable tasks that do not involve imagery.

[13] A possible control for this explanation would be to study patients whose loss of peripheral vision and delay in testing followed roughly the same pattern as Farah’s patient but in which the damage was purely retinal. The expectation is that under the same instructional conditions such patients would also exhibit tunnel imagery, even though there was presumably no relevant cortical damage involved. Another control would be to test the patient immediately after surgery, before she learned how the visual world looked to her in her post-surgical condition.

[14] That a topographic display is involved in vision is hardly surprising, since we know
that vision begins with retinal images. But before either retinal or cortical patterns become available to cognition as percepts, they have to be interpreted; this is what vision is for, it is not for turning one retinotopic pattern into another. The original motivation for hypothesizing a visual image was to account for the completeness and spatially extended nature of visual perception despite the incompleteness of retinal information, the stability of the perceptual world despite eye movements, and the robustness of recognition despite differences in the size and location of objects on the retina (these are among the reasons Kosslyn 1994, gives for needing a display that also serves as the screen for mental images). It is now pretty clear that there are no visual images serving these purposes in vision (see, for example, O'Regan & Noë 2002).

[15] This paper has been concerned primarily with the picture-theory account of imagery because it appears to be the overwhelmingly dominant one. I have not attempted to review other options, such as those proposed by (Barsalou 1999b; Thomas 1999). In any case these other approaches are not claimed to provide an explanation of the many experimental findings sketched in this paper.

[16] The claim that images are epiphenomenal or have no causal role rests on an ambiguity about what one means by an image. As Block has correctly pointed out (Block 1981), the appeal to epiphenomenalism is either just another way of stating the disagreement about the nature of a mental image or else it simply confuses the functional or theoretical construct “mental image” with the experience of having a mental image.