

SEEING AND VISUALIZING: IT'S NOT WHAT YOU THINK An Essay On Vision and Visual Imagination*

{ CHAPTERS 6, 7, 8 FOR CLASS USE }

ZENON PYLYSHYN, RUTGERS CENTER FOR COGNITIVE SCIENCE

Table of Contents

6. Seeing With the <i>Mind's Eye 1: The Puzzle of Mental Imagery</i>	6-1
6.1 What is the puzzle about mental imagery?	6-1
6.2 Content, form and substance of representations	6-6
6.3 What is responsible for the pattern of results obtained in imagery studies?	6-8
6.3.1 Cognitive architecture or tacit knowledge	6-8
6.3.2 Problem-solving by "mental simulation": Some additional examples	6-13
6.3.2.1 Scanning mental images	6-13
6.3.2.2 The "size" of mental images	6-17
6.3.2.3 Mental "paper folding"	6-19
6.3.2.4 Mental Rotation	6-21
6.3.3 A note concerning <i>cognitive penetrability</i> and the appeal to tacit knowledge	6-24
6.3.4 Summary of some possible reasons for observed patterns of imagery findings	6-26
6.4 Some alleged properties of images	6-28
6.4.1 Depiction and mandatory properties of representations	6-28
6.5 Mental imagery and visual perception	6-31
6.5.1 Interference between imaging and visual perception	6-33
6.5.2 Visual illusions induced by superimposing mental images	6-33
6.5.3 Imagined versus perceived motion	6-35
6.5.4 Extracting novel information from images: Visual (re)perception or inference?	6-37
6.5.5 What about the <i>experience</i> of visualizing?	6-41
7. Seeing With the <i>Mind's Eye 2: Searching for a Spatial Display in the Brain</i>	7-1
7.1 Real and "functional" space	7-1
7.2 Why do we think that images are spatial?	7-6
7.2.1 Physical properties of mental states: crossing levels of explanation	7-6
7.3 Inheritance of spatial properties of images from perceived space	7-10
7.3.1 Scanning when no surface is visible	7-12
7.3.2 The exploitation of proprioceptive or motor space	7-13
7.4 The search for a real spatial display	7-17
7.4.1 Aside: Does biological evidence have a privileged status in this argument?	7-18
7.4.2 The argument from differential activity in the brain	7-19
7.4.3 The argument from clinical cases of brain damage	7-23
7.5 What would it mean if all the neurophysiological claims turned out to be true?	7-25

* Manuscript of a forthcoming book from MIT Press. Please do not quote without permission. © 1998 by Zenon Pylyshyn

7.5.1	The ‘mind’s eye’ must be very different from a real eye	7-26
7.5.2	The capacity for imagery is independent of the capacity for vision.....	7-27
7.5.3	Images are not two-dimensional displays	7-28
7.5.4	Images are not retinotopic	7-29
7.5.5	Images do not provide inputs to the visuomotor system.....	7-30
7.5.6	Examining a mental image is very different from perceiving a display.....	7-31
7.5.7	What has neuroscience evidence done for the “imagery debate”?.....	7-35
7.6	<i>What, if anything, is special about mental imagery?</i>	7-36
7.6.1	What I am <i>not</i> claiming: Some misconceptions about objections to the picture-theory	7-36
7.6.2	What constraints should be met by a theory of mental imagery?.....	7-39
8.	Seeing With the Mind’s Eye 3: Visual Thinking	8-1
8.1	<i>Different “styles” of thinking</i>	8-1
8.2	<i>Form and content of thoughts: What we think with and what we think about</i>	8-1
8.2.1	The illusion that we experience the <i>form</i> of our thoughts	8-1
8.2.2	Do we think in words?.....	8-2
8.2.3	Do we think in pictures?	8-4
8.2.4	What form must thoughts have?.....	8-6
8.3	<i>How can visual displays help us to reason?</i>	8-7
8.3.1	Diagrams as logical systems that exploit visual operations	8-7
8.3.2	Diagrams as guides for derivational milestones (lemmas).....	8-10
8.3.3	Diagrams as a way of exploiting visual generalization	8-11
8.3.4	Diagrams as ways of tracking instances and alternatives	8-15
8.3.5	Diagrams as non-metric spatial models and spatial memory	8-16
8.3.6	Diagrams drawn from memory can allow you to make explicit what you knew implicitly	8-17
8.4	<i>Thinking with mental diagrams</i>	8-18
8.4.1	Using mental images	8-18
8.4.2	Using mental models	8-20
8.4.3	What happens during visualization?	8-21
8.5	<i>Imagery and imagination</i>	8-22
8.5.1	How does creativity connect with imagery?.....	8-25
8.5.2	Enhancing creative thinking.....	8-26
References	8-28

6. SEEING WITH THE *MIND'S EYE* 1: THE PUZZLE OF MENTAL IMAGERY

6.1 What is the puzzle about mental imagery?

In earlier chapters I discussed various connections among the world, the visual system, and the central cognitive system that is responsible for reasoning, inference, decision-making and other rational processes. In the course of this analysis we have found much that is counter-intuitive about how the mind is organized and how it connects with and represents the world. Yet nowhere does our intuition go astray more than when we consider what happens when we recall a past event by imagining it or when we reason by imagining a situation unfold before our “mind’s eye.” Our introspection is very persuasive here and it tells us quite unequivocally that when we imagine something we are in some important sense *seeing* it and that when we solve a problem by imagining a situation unfold in our “mind’s eye” we have but to pay attention and notice what happens. No intervention on the part of reasoning appears to be involved in much of this process.

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 their imagination. 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 we can predict the dynamics of physical processes that are beyond our ability to solve analytically. In fact they may even be beyond anyone’s ability to solve analytically. Consider the behavior of a coin that has been spun on its edge. As it topples it rolls around faster and faster until, with a final shudder, it lies still. The behavior of this coin, which accelerates as it loses energy, has only recently been mathematically solved (Moffatt, 2000). Yet many people can imagine the behavior in question when asked to visualize the situation. Why? Is it because something in their imaging mechanism inherently obeys the relevant laws? Or is it perhaps because, under the right circumstances, they recall having seen a dropped coin behave this way? In either case, the intuition that the behavior of one’s image is automatic is very strong. There seems to be something involuntary and intuitive about how the action unfolds in one’s image; the impression that it is properties of your image, rather than any involvement of reasoning, which is responsible for the action that you passively observe with your mind’s eye and your visual system, is similarly very strong. Can this possibly be true?

Imagine that you climb a ladder and drop a ball from a ten-foot height. Observe it in your mind’s eye as it falls to the ground. Now do the same while imagining that you drop it from a five foot height. Does it not take longer to fall to the ground from a height of ten feet than from a height of, say, five feet? How much longer, do you think? You could actually measure the time it took for the ball to drop in your imagined scenario and get an exact answer which would surely confirm that it takes longer to drop from a greater height. You could even plot the time as a function of distance and perhaps even as a function of imagined weight (and you would probably find, as Ian Howard did, that the time was proportional both to the height and to the weight – unlike what would happen if you actually dropped a real ball¹). Now imagine that you are riding a bicycle along a level road, peddling as hard as you can. Then imagine you have come to a hill. Does

¹ In a series of unpublished studies, Ian Howard of York University, showed that people’s naïve physics, as measured by their predictions of falling objects, conformed to Aristotelian rather than Newtonian or Galilean principles. Howard dropped metal rings of various sizes behind an opaque screen and asked subjects to catch them by poking a rod through one of several carefully located holes in the screen provided for the purpose. The measured times showed that subjects assumed a constant velocity, rather than the correct constant acceleration, and that the assumed velocity increased with the weight of the ring, contrary to Galilean principles (i.e., that objects fall with constant acceleration, rather than constant velocity, and the acceleration is independent of mass).

the bicycle not slow down without you having to think about it? What about when you come to the downhill side of the hill? You can even stop peddling completely and you will probably find that the bicycle in your image continues to speed down the hill.

The illusion of the autonomy of the imagined scene is even more persuasive in the case of purely visual or geometrical properties. Imagine a very small mouse in the far corner of the room. Can you easily see whether it has whiskers? Now imagine that you are very close to it (perhaps it is in your hand). Isn't it much easier now to see whether it has whiskers? Close your eyes and imagine you are looking at the bottom left corner of the wall in front of you. Now imagine that you shift your gaze to the upper right corner. Did your gaze move through the intermediate points as you shifted from one corner to the other? Did you notice any of the things on the wall as you shifted your gaze? Do you think it would have taken you longer to shift your gaze if the region you were imagining had been smaller (say confined to just a picture on the wall)? Imagine a rectangle and imagine drawing the diagonal from the top left corner to the bottom right corner. Now imagine drawing another line, from the bottom left corner this time, to the middle of the right-hand side. Does this second line cross the first one you drew? And if so, does it cross it below or above its midpoint? Did you have to do any geometrical analysis to answer that question, or did you just read it off your image? Imagine an upper case letter **D**. Imagine it rotated counterclockwise by 90 degrees. Now imagine the letter **J** attached to it from below. What is the shape of the resulting combined figure? It seems as though you can do this entirely in your mind's eye and "notice" what the combination of the two figures "looks like" an umbrella without thinking about it. Such examples seem to show that mental imagery has a "life of its own" and unfolds without your rational intervention. They also seem to show that imagery makes use of the process of *visual* recognition. But does it? Notwithstanding one's strong intuitions about all these examples, there is much more to what your mental image does and what it "looks like" than meets the eye – even the "mind's eye". As we saw in Chapter 1, the more familiar and vivid our experience of our inner life is, the less we can trust our impressions to provide the basis for a scientific theory of the underlying causes. I will return to more examples of this later.

Opposing the intuition that your image has an autonomous existence, and unfolds according to its own laws and principles, is the obvious fact that it is *you* alone who controls your image (of course the same might be said for other forms of conscious thought, such as inner dialogue, which are subject to the same erroneous intuitions, as we will see in chapter 8). Perhaps, as (Humphrey, 1951) once put it, the assumption that the image is responsible for what happens in your imagining puts the cart before the horse. It is more likely that the image unfolds as it does because you, the image creator *made* it do so. For example, does it not seem plausible that the way things unfold in your image is actually guided by what you know would happen in real life? You can imagine things being pretty much any size, color or shape that you choose and you can imagine them moving any way you want. 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 baseball trajectory. You can imagine all sorts of impossible things happening – and cartoon animators frequently do, to everyone's amusement. The wily coyote chasing the roadrunner runs onto what turns out to be a picture of the road or sails off the edge of the cliff but does not fall until he sees that there is no support under him. Or the roadrunner sees the coyote approaching and suddenly appears behind his back. We can easily imagine the laws of physics being violated. In fact unless we learned certain Newtonian principles of physics we do not even correctly imagine what happens in certain real situations – such as in the earlier example of the time course of an object falling to earth (as mentioned in note 54).

If we can imagine anything we like, why then do we imagine the things we do in certain circumstances? Do we imagine things happening in a certain way merely because we *know* that this is how they would happen? What if we can't predict how they will happen without going through the process of imagining them; does this mean that imagery contains different kinds of "knowledge" than other forms of representations? And what exactly would that mean? Is it that imagery involves the use of *different* sorts of reasoning mechanisms? It may feel as though we can imagine pretty well anything we like, but is that so? Are there events, properties or situations that we cannot imagine, and if so, why not? While we can imagine the laws of physics being violated, can we imagine the axioms of geometry 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 (in which, for example, the diagonal of a right-angled triangle was longer than the sum of the other two sides). Before concluding that such examples illustrate the intrinsic geometry of images, however, 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. The answer is by no means obvious. For example, do you know where various edges, shadows and other contours would fall in a four-dimensional or non-Euclidean space? If you don't know how something would look, then how could you possibly have an image of it, since having an image of something *means* to imagine how it *looks*. 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 that images have the sort of inherent geometrical properties that we associate with pictures.

We also need to keep in mind that despite one's intuitions about such examples, there is reason to be skeptical about what one's subjective experience reveals about the nature of the mental image (both its content and its form). After all, when we look at an actual scene we have the unmistakable subjective impression that we are examining a detailed panoramic view, yet as we saw in Chapter 1, there is now considerable evidence that we encode and store very little in a visual scene unless we explicitly attend to the items in question, which we do only if our attention or our gaze is attracted to it (Henderson & Hollingworth, 1999). The natural implication of our phenomenology appears to be quite wrong in this case, even though the phenomenology itself is not to blame; things appear to us the way they appear – it's how we interpret this appearance that is in error. The information we have about a scene is not in the form of a global picture, for reasons discussed in chapter 1, section 1.4 (e.g., because non-retinal information appears to be much more abstract and conceptual than retinal information, because information from successive glances cannot be superimposed, and so on). 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.

What needs to be kept in mind is that the content of our mental images, both the explicit information it contains and the dynamics of how that changes, is the joint consequence of (a) what we intend our mental image to show, (b) what we know about how things in the world look and how they tend to unfold in time, and (c) the way our mind (or perhaps only the part of it that specializes in mental imagery) constrains us (there are some other possible extrinsic reasons as well – see Section 6.3.4). Discovering the boundary between the major determiners of image content (in particular, between what we *know* or *intend* and the constraints imposed by the mechanisms or the particular form of representation used – in other words, by the way our brain is structured) is the central problem in the study of mental imagery. Both the impression that we can imagine whatever we please and the impression that our images have a life of their own are illusions. Our task as scientists is to try to steer our way between the Charybdis of total plasticity and the Scylla of the autonomous unfolding. This task is no different from the one faced by cognitive scientists in every area of

cognition since what we can and do think is similarly a joint consequence of the plasticity of thought (we believe we can think any thought there is to think) and of the constraints imposed on them (on what we can represent, conceptualize, or infer) by the nature of mind. The latter constraints arise from what we call the *cognitive architecture* and we will have more to say about this concept later. In the case of the particular cognitive activity that is accompanied by vivid experiences, such as the experience of seeing an image in our mind's eye or experiencing ourselves thinking in sentences of our language, the temptation to reify the experiential content into a theory of the form of our representations seems very nearly inescapable, leading to the ubiquitous view that the mind has two distinct modes of thought: linguistic and imagistic (or pictorial). In fact these two ways of experiencing our thoughts neither exhaust the ways in which we can think nor do they provide a useful characterization of what thinking consists of – but more on this in the Chapter 8. It is precisely the power of introspection to, on the one hand, provide a window into what we are thinking *about* and, on the other hand, mislead us into believing that we can see the *form* in which our thoughts are encoded and the nature of the thinking process itself, that creates difficulty in coming to an understanding of the nature of perception and thought. It is the main reason why there is a ubiquitous problem of grasping what a scientific theory of reasoning with mental images might be like.

The view to which we are tempted by our introspection is the view discussed earlier in Chapter 1 (it is what Dan Dennett has called the “Cartesian Theater” view of the mind, Dennett, 1991). It is the view that when we think using mental images we are actually creating a picture in our mind. And when we reason in words we create sentences in our mind. In both cases there is the concomitant assumption that someone (or something) perceives these pictures or sentences. Henceforth I will refer to such views (at least as applied to the visual modality) as the “picture theory” of mental imagery. This is sometimes contrasted with an alternative that posits symbol structures, of the sort that appear in artificial intelligence or other computational models. As we will see in chapter 8, there are very good reasons for believing that thought takes place not in natural language or images, but in what has sometimes been called the “language of thought” (LOT) or *lingua mentis* (Fodor, 1975). For present purposes I will not be concerned with the LOT theory of the mental structures that underlie the experience of mental imagery. Rather I will be concerned to show that the intuitively appealing *picture theory* is implausible. I will do so by arguing that none of the phenomena and experiments distinguishes between a picture theory and the class of symbolic or “language of thought” possibilities. Because the latter are needed to explain non-imaginal thinking and reasoning, I will take the view that thoughts have the same form regardless of how they are experienced as our *null hypothesis*. It should be understood, however, that the *null hypothesis* in this case, as in all cases of empirical inquiry, is held provisionally subject to rejection by adequate disconfirming evidence or a more persuasive theory. The question then, is whether any of the arguments or empirical evidence presented to date suggests that we should reject this *null hypothesis*.

But why should we shun the intuitively appealing view in favor of something so counter-intuitive? In chapter 1 I alluded to some of the reasons why in the case of visual perception, the intuitive view is inadequate and incapable of accounting for the empirical facts. Before elaborating this argument and extending it to cover the case of mental imagery, let us stand back and try to get a broader picture of the problem of mental imagery that needs explaining. Before doing so, however, I do want to clarify one point that is often misunderstood. Contrary to what many critics have assumed, I do not claim that images do not really exist or are merely “epiphenomenal” (Shepard, 1978b). The notion of something being an epiphenomenon is itself misleading. Usually what people mean when they accuse me of making this claim is that they do not agree my view (as Block, 1981b, has correctly pointed out). There can be no question of whether the experience of imagery exists, nor is there even much disagreement about its phenomenal

character. The scientific problem is the explanation of the causal events that underlie this phenomenal experience. In cognitive science, these causal events typically take the form of information processing operations performed on some representations. The explanation of what underlies this process need not, and in general will not, appeal to how they are experienced – to the fact they appear to us to be like pictures of the scene we are imagining. This is very much like saying that the way physical objects look to us is not part of our theory of how they take part in physical and chemical reactions. For that we need a different theory, a theory of their underlying causal structure, which always adverts to invisible things and invisible properties and forces.

Of course the phenomenology is also interesting and one may be able to tell a fascinating story about how things appear. In fact (Shepard, 1978a) has done a remarkable job of illustrating many of his hypnagogic (near-sleep), entoptic (externally caused, though not by light), and dream images. These are fascinating accounts, but it is not clear how they should be taken within a causal theory. In commenting on a paper of mine, (Dalla Barba, Rosenthal, & Visetti, 2002) remark that phenomenology was never intended to provide a causal explanation. What it does is provide an account of how things are experienced, and in so doing it may show how some experiences are like others – it may provide a taxonomy of how things appear. On the basis of their (and other writers') phenomenal experience of mental imagery and of perception, Dalla Barba et al. conclude that images are pictorial but that they also differ significantly from the experience of vision. Whether phenomenally based taxonomies can be accommodated in an information-processing theory, or whether they can help to decide among alternative theories, is at present an open question (both in the study of mental imagery and in the study of vision).

The most important idea that must guide us in trying to understand the nature of mental imagery is the question of which properties and mechanisms are *intrinsic* or *constitutive* of having and using mental images, and which arise because of what we believe, intend, or otherwise attribute to *that which we are imagining*. The central question we need to ask is which aspects of “visual” or imaginal (or imagistic) thinking occurs because of the special nature of the imagery system, rather than because of the nature of thinking in general, together with our tacit knowledge of the situation being imagined, how this knowledge is organized, and how we interpret the imagining task. We also need to ask whether what is special about imaginal thinking might simply be the fact that mental imagery is associated with certain contents or subject matter, such as the appearances of the things we are thinking about, and not on the way the information is encoded or on any special mechanisms used in processing it. Notice that it *could* be the case that when we think about spatial layouts, or about the appearance of objects (e.g., their shapes and colors) or other visual properties that tend to elicit the experience of mental imagery, we find that certain psychophysical phenomena are observed due to the fact that we are then thinking about concrete visible properties rather than abstract properties. This would be no different from finding that certain properties (like differences in reaction times to different words or the ability of different secondary tasks to interfere with the thinking) are observed when we think about economics or music or sports. It is plausible that when you are solving problems in economics your performance is degraded by secondary tasks involving economic concepts, the way that thinking about spatial topics is disrupted by tasks which themselves involve spatial concepts. Differences attributable to the *topic* of thinking are clearly not what those who postulate a separate image system have in mind. If what goes on during episodes of thinking with images is just the sort of thing that goes on when imagery is not being used, then the postulation of a special image system would be redundant and gratuitous, *regardless of whether we have a worked-out theory of any kind of reasoning*.

A number of important distinctions have to be made before evidence from hundreds of experiments

dealing with mental imagery can be interpreted. As I suggested above, a serious possibility is that the experiments that ask subjects to solve a problem or answer a question by using their mental image may be interpreted by the subject as the request to say what would happen if they were to *see* the corresponding imagined event taking place. But this deflationary explanation requires that we distinguish between explicit and tacit knowledge, between patterns of reasoning that arise from habit or preference, as opposed to patterns that arise because of the nature of the fixed mechanisms involved, and between patterns that are intrinsic to reasoning with images as opposed to those that would arise no matter what form or modality of reasoning was being used. Even if we do find that there are properties intrinsic to reasoning with images, we must still ask whether these arise because of certain imagery-specific forms of representation or processing that are being deployed or, as generally claimed, because imagery involves applying specifically *visual* mechanisms to perceive an inner picture-like (depictive) pattern.

6.2 Content, form and substance of representations

One thing that everyone agrees on is that mental images are *representations*: they encode information about the world. In this and subsequent chapters we will be concerned with visual images that primarily, though not exclusively, encode the visible world. When we speak of a representation, there are at least three levels of analysis at which we can theorize² (for a detailed discussion of these “levels” see Pylyshyn, 1984a). At the *first* level we can ask about the content, or what the representation *represents* – what it is *about*. The thing represented need not even exist (e.g., we can represent unicorns with no difficulty), and even if it does exist, the content of the representation is different from the thing in the world being represented. That’s because the content of the representation consists not only in some particular thing in the world, but also the description under which it is represented or what it is represented *as*. If we represent a certain physical thing as Dr. Jeckyl our representation has a different content than if we represent the very same physical thing as Mr. Hyde. If we represent a certain point of light as the planet Venus, the representation has a different content than if we represent it as the morning star or as the evening star, even though all these representations refer to the very same physical body. The difference in how it is represented (or what it is represented *as*) has real consequences for potential behavior. A great deal has been said about the notion of content in philosophy, but for our purpose all we need to acknowledge is that there is more to being a representation than having a certain form.³ For many purposes we also need to talk about the content of the representation and we need to distinguish between properties of the representation’s form and properties of the representation’s content.

At the *second* level of analysis, we can inquire about the *form* of the representation, the system of codes by which mental objects can represent aspects of the world. These codes need not be discrete symbols, but they do have to embody some principles by which they combine and by virtue of which they are able to represent novel things – they need to form a productive system (see Fodor & Pylyshyn, 1988). In the case of language, these principles are referred to as the *syntax* of the language. Differences in representational content (in what the linguistic objects represent) arise from corresponding differences in the form of the representation, which means the terms (codes) that appear and the way they are syntactically structured. We will devote considerable attention in this and the next chapter to the question: What is the *form* of mental images? It is at this level that we can raise such questions as whether mental images use a different system of encoding from other forms of thought. The formal or symbol level divides into two kinds of questions: (a)

² (Newell, 1980) refers to these levels as the knowledge level, the symbol level, and the physical level, respectively. He also mentions additional levels that are important in actual digital computers (e.g., the register transfer level). There may, in fact, be reason to include other levels in the case of mental processes (e.g., an abstract neurological level that accounts for additional generalizations) but as far as I am aware none has been proposed so far.

³ This quick sketch does not do justice to this complex subject. For example, the content of a representation also includes certain logical components (given by logical constants like AND, OR, and NOT) as well as quantifiers (SOME, ALL, EVERY, EXISTS), modalities (MUST, CAN, SHOULD), and propositional attitudes (BELIEVES, KNOWS, FEARS, HOPES).

What are the (relatively) fixed computational resources out of which processes are composed, and which determine the form that representations may take, and how are they organized (e.g., are they encapsulated, like early vision, or can they communicate freely among themselves)? (b) What is the forms or structure of images (e.g., are they two-dimensional displays, as some have claimed?) and what are the particular processes that underlie reasoning with mental images (e.g., are they visual)?

The *third* level of analysis of mental representations is concerned with how representations are realized in biological tissue or implemented in hardware. In chapter 7 we will consider evidence at this third level of analysis. We will see that the distinction between the information-processing level and the biological or physical level often gets blurred because both are concerned with the question: *How is the function realized?* These two levels often even use the same terminology – for example, they may refer to the “size” of mental images, where a systematic ambiguity is retained between some formal property of images that may be responsible for certain observed experimental phenomena and a literal physical property of its instantiation in brain tissue. We can ask the question about how some process is carried out or how something works in at least two senses or at two different levels of abstraction. One level considers functional or computational mechanisms in information-processing terms, while the other focuses on brain mechanisms described in neuro-anatomical terms. Unfortunately, knowledge of neural mechanisms is not sufficiently advanced to allow this level of analysis to elucidate the computations or algorithms are involved,⁴ so the most that we typically get from this level is a general *taxonomy of functions* derived from brain damage and neural imaging data. The other level at which we can address the question of how something is done in the brain is by asking what basic operations and formal properties the brain makes available. This question then really belongs to the second (formal) level, though it is often confused with the biological level. One of the discoveries of the last 40 or so years of information processing psychology is that it is possible to ask the question: *How is it done?* without requiring an answer in terms of biology or physics. To ask how it is done in this more abstract sense is to ask how it is realized by information processing resources (basic operations, forms of encoding, types of memory, and so on) that constitute what we call the *cognitive architecture*. The cognitive architecture is just a more abstract description of the way the brain or computer functions in relation to its capacity to process representations, but it is a level of description that captures the essential aspect of information processing abstracted away from the sorts of details that differ from occasion to occasion and from person to person.

This brings us to the notion of cognitive architecture, as a description of the relatively fixed mechanisms and properties of the brain, couched in information processing terms. This notion is absolutely central to theorizing about cognitive processes. Elsewhere I have written extensively on the topic of cognitive architecture (Pylyshyn, 1984a, 1991a, 1996). For present purposes I wish only to point out that the distinction between cognitive architecture and knowledge-based processes arises in the present context because they involve different types of explanatory principles, which are often confused in discussions about mental imagery. For example, we have already seen how mental images are often discussed as having a certain form or having certain special properties (e.g., size, distance), when what is meant is that the *content* of images (or the things referred to or depicted in the images) have these properties. The property of being larger or smaller or been further or closer are among such properties. There is clearly a difference between claiming that you have an image of something big (or red or heavy or slow or whatever) and claiming that the image itself is big (or red or heavy or slow, and so on).

While there is general agreement that we need all three levels of description when we discuss systems of representation it has not always been the case that these levels have been kept distinct. In particular, when

⁴ Indeed, Randy Gallistel (Gallistel, 1994) has pointed out that theorizing about how the brain may carry out certain basic functions is currently stuck in a psychology that was discredited decades ago – a psychology that takes association to be the basic operation, rather than the manipulation of symbols (including real-valued magnitudes processed in a register architecture).

one claims certain properties for images it is important to be clear as to the level at which our claims apply. In Chapter 7 we will see that it is especially true when we consider claims such as that images *have* or *preserve* metrical or spatial properties of the world they represent.

6.3 What is responsible for the pattern of results obtained in imagery studies?

6.3.1 Cognitive architecture or tacit knowledge

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, preferences, habits, or interpretation of the task at hand, is central, not only to understanding the nature of mental imagery, but to understanding mental processes in general. The former sorts of effects (those attributable to the intrinsic nature of mechanisms) invoke what has been called the cognitive architecture (Fodor & Pylyshyn, 1988; Newell, 1990; Pylyshyn, 1980, 1984a, 1991a, 1996) – one of the most important ideas in cognitive science. Cognitive architecture 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 nor any other representations (e.g., fears, hopes, fantasies, etc). In other words when you form an interpretation of a certain task, or find out new things about a situation you are thinking about, or when you draw inferences from what you know, or weigh the options and make a decision, your cognitive architecture does not change. Of course, if as a result of your beliefs 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, 1984a, Chapter 7). This informal characterization and the following example will have to do for present purposes.

To make this point in a more concrete way, I invented a somewhat frivolous but revealing example, involving a certain mystery box of unknown construction whose pattern of behavior has been assiduously recorded (Pylyshyn, 1984a). This box is known to emit long and short pulses with a reliable recurring pattern. The pattern (illustrated in Figure 6-1) can be described as follows: pairs of short pulses usually precede single short pulses, except when a pair of long-short pulses occurs first. In this example it turns out that the observed regularity, 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. These two sorts of "reasons" for the observed pattern (intrinsic or extrinsic) are analogous to the architecture versus tacit knowledge distinction and is crucial to understanding why the box works the way it does, as well as to why certain patterns of cognition occur.

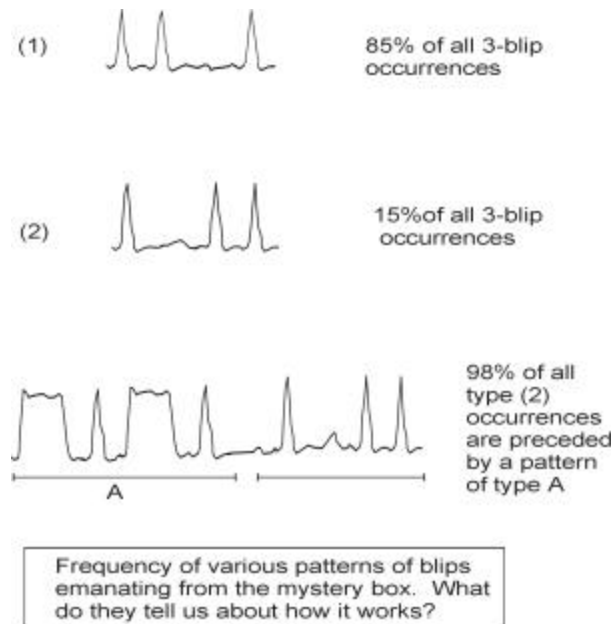


Figure 6-1. Pattern of blips observed from a box in its typical mode of operation. The question is: Why does it exhibit this pattern of behavior? What does this behavior tell us about how it works?

The reason why this particular pattern of behavior occurs in this case can only be appreciated if we know that the pulses are codes, and the pattern is due to a pattern 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 pattern in the way English words are spelled (the principle being that generally *i* comes before *e* except after *c*). Similarly, I have argued that very many of the patterns observed in mental image research reflects a principle that subjects believe holds in the imagined world, and not a principle of their mental architecture. The pattern arise from the fact that subjects know what would happen if they were to see certain things unfold before their eyes, and they make the same thing happen in their imagined simulation. The reason that the behavior of both the mystery code box and the cognitive system do not reveal properties of its intrinsic nature (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 nature (their “wiring” or their causal structure) in order to change their behavior. The way the behavior can be altered provides the key to how you can tell what is responsible for the observed regularity. This is the basis for the methodological criterion called “cognitive penetrability” to be described in section 6.3.3.

Clearly it is important to distinguish between architectural and knowledge-based explanations in understanding mental imagery. I noted earlier that in order to understand what goes on in mental imagery it is essential to distinguish the case where (a) people are merely making the mental representation underlying their phenomenal image (*whatever* that turns out to correspond to in our theory) have the contents that they independently believe would be seen or would occur in certain situations, from the case where (b) it is the very fact of putting their thoughts in the particular form of representation corresponding to a mental image – and thus being constrained by the properties of this form of representation – that is responsible for the outcome. In case (b) the claim is that properties intrinsic to images, to their form or their particular

realization in the brain, that results in thoughts unfolding the way they do. In other words, in this second case we are claiming that the observed properties of the image-based thoughts are a direct consequence of properties of the special cognitive architecture used in mental imagery. In case (a) where people simply make their image do what they believe would happen if they were seeing the event happen, the properties of the image representation are irrelevant to explaining the way the thoughts unfold: nothing is gained by postulating that an image representation has certain properties since these properties do not figure in any explanation and they do not in any way constrain the outcome. Thus saying that an image is like a picture may reflect the phenomenology of imagery, yet if case (a) were the correct analysis of what is going on, the pictorial format would be theoretically irrelevant to explaining the outcome of thinking with images. Thus the phenomenology-based theory of the representation underlying a mental image would be doing no work because the real explanation would lie elsewhere, for example it would lie in what people decided to put into their phenomenal image or what they made it do.

To see that the distinction between knowledge-based and architecture-based accounts of why things work the way they do really does make a difference to our understanding of how imagery works, try to imagine a physical situation whose operating principle is completely unknown to you. For example, imagine that you have a jar filled with sugar and a jar filled with water. Imagine, in your mind's eye, that the water is slowly poured into the jar of sugar, as shown in Figure 6-2. Does the water in the sugar-filled jar begin to overflow – and if so at what point in the pouring does it do so? In this case it seems clear that what will happen in your imagination will depend on what you know (or believe) would happen in the world if you observed such an experiment being performed. Your imagination clearly does not embody the subtle principles by which solids dissolve in fluids, which involves understanding how molecules of certain solids can take up the spaces between molecules of the fluid. What happens in your imagination is just exactly what you *think* would happen (perhaps based on what you once saw happen), nothing more. Someone who claimed that it was up to their image to determine what will happen and that it was the properties of their imagery system that generated the result, would be letting their phenomenal experience cloud their judgment. To see that this must be so, try making your image do something different by just willing it to!



Figure 6-2 Imagine pouring water into a beaker full of sugar. Does it eventually overflow?

Take another example not involving such an obscure principle of physics. Ask yourself what color you see if you look at a white wall through a yellow filter and then gradually superimpose a blue filter over the yellow 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 would look. Try this out on yourself. Imagine looking at a white wall through a blue filter and a yellow filter and then bring them into overlapping positions, as illustrated (without the benefit of color) in Figure 6-3. What color do you “see” in your mind's eye in the overlap region? More important, ask yourself why do 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 the imagery “medium”, from the intrinsic character of the color encoding and display mechanism deployed in imagery in images, just as the parallel case of visual color mixing arises from the intrinsic character of the color receptors in the eye, together with the character of light that is transmitted through colored filters. But since there can be no doubt that you can make the color of the overlap portion of the filters in your mental image 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? This is where the notion of *tacit knowledge*⁵ plays an important role in cognitive science (Fodor, 1968, see also section 6.3.3). It seems clear in this case that the color you “see” depends on your tacit knowledge either of principles of color mixing or of these particular color combinations (having seen something like them in the past). In fact people who do not know about *subtractive* color mixing generally get the above example wrong; mixing yellow light with blue light in the right proportions produces white light, but overlapping yellow and blue filters leads to green light being transmitted.

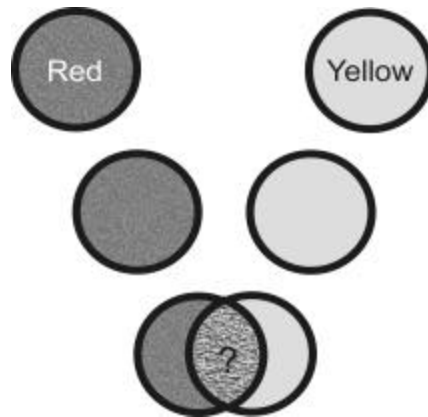


Figure 6-3. Imagine the red and yellow disks moving closer and closer until they overlap. What color do you see in your image where the two disks overlap?

When asked to do this exercise, some people simply report that they see no color at the intersection, or a washed-out indefinite color. Still others claim that they “see” a color different from the one they report when asked to answer without imagining the filter scenario (as reported in Kosslyn, 1981). Cases such as the latter have made people skeptical of the tacit knowledge explanation. There are indeed many cases where people report a different result when using mental imagery than when they are asked merely to answer the question without using their image. Appeal to tacit (inexplicit) knowledge may be crucial, but that does not mean one can solicit the relevant tacit knowledge by merely asking a subject. It is a general property of reasoning that the way the question is put and the reasoning sequence used to get to the answer can affect the outcome. As I will suggest in section 6.3.2.3, knowledge can be organized in many different ways and it can also be accessed in many different ways – or not accessed at all if it seems like more work than it is worth.

To illustrate this point, consider the following analog of the color-mixing task. Imagine your third grade teacher writing the following on a blackboard: “ $759 + 356 = \underline{\quad}$?” Now, as quickly as you can, without stopping to think about it, imagine that the teacher continues writing on the board. What number can you “see” the teacher write in the blank? Now ask yourself why you saw *that* number rather than some other number being written in the blank. People will imagine different things in this case depending on whether they

⁵ Outside of epistemology, people typically do not distinguish between “belief” and “knowledge” so that it may turn out that some “knowledge” might be false or at any rate not justified. I use the term “knowledge” simply because it has been in general use in psychology and in artificial intelligence and the distinction between belief and knowledge is immaterial in the present context.

believe that they are supposed to work it out or whether in the interest of speed they should 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 (though it needn't be) a different task from the task of solving a certain intellectual puzzle about colors, as you can see if you consider the difference between the various ways of you might go about filling the blank in the arithmetic example. The difference can be like the difference between free-associating to a word and giving its definition. The task of imagining something unfolding in your "mind's eye" is a special task: It's the task of simulating as many aspects of the visual situation as possible, as many aspects as you can – not because you are being led on by the experimenter and not because of some special property of mental imagery, but *because this is what it means to "imagine X happening."*

In most of the cases studied in imagery research, it would be odd if the results did not come out the way picture theories would predict, and if they did, 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 construct a vivid and detailed auditory image of a competent performance of the Minute Waltz, played on a piano in front of you, the failure of the imagined event to take approximately one minute would simply confirm that you had not carried out the task properly (or at all). Taking roughly one minute is inherent in a real performance and thus it is natural to assume it to be indicative of a good imagined re-creation or simulation of such a performance. To realistically imagine a musical performance of a piece *means* to imagine (i.e., think of) each token note being played in the right order and at the right loudness and duration, whether or not it entails that certain sensory qualities are "perceived" by the "mind's ear." In other words, regardless of what else you believe goes on when you imagine hearing the piano piece being played, one thing that the task requires is reproducing a sequence of mental states or thoughts corresponding to "hearing" the sequence of notes in the right order and at roughly the right durations. Thus in this case, regardless of the form of representation involved and regardless of what mental processes take place during such episodes of imagining, they have no bearing on the observed outcome (i.e., the time taken) because the outcome is attributable to the some of the imager's tacit knowledge about the Minute Waltz.

Finally, let me emphasize again what kind of tacit knowledge is relevant to this discussion, because there has been serious misunderstanding of this question. The only knowledge that is relevant to the tacit knowledge explanation is knowledge of *what things would look like* in certain situations, in particular in situations like the ones in which subjects in mental imagery experiments are to imagine themselves. Thus it is not a criticism of this type of explanation to point out, (as Farah, 1988, pp 314-315 does), that people are unlikely to know how their visual system or the visual brain works. That's not the tacit knowledge that is relevant. Nor is it the tacit knowledge of what results the experimenter expects (sometimes referred to as "experimenter demand effects") as many have assumed (Finke & Kurtzman, 1981b). Of course, the latter is highly relevant as well, and may even explain some of the findings in imagery studies (as suggested by Banks, 1981; Intons-Peterson, 1983; Intons-Peterson & White, 1981; Mitchell & Richman, 1980; Reed, Hock, & Lockhead, 1983; Richman, Mitchell, & Reznick, 1979), but it is not what I mean by the tacit knowledge explanation. All I mean is that subjects in studies where they are asked to "imagine X" use their knowledge of what "seeing X" would be like to simulate as many of these effects as they can. Doing this successfully, of course, depends on having certain psychophysical skills, such as the ability to generate time intervals proportional to certain computed magnitudes (Fraisse, 1963), or to compute the time-to-collision of moving objects (as we will see section 6.5.3).

6.3.2 Problem-solving by “mental simulation”: Some additional examples

The idea that what happens in certain kinds of problem solving can be viewed as *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). Take, for example, the question of how we manage (rather successfully) to predict other people’s behavior in everyday life. One proposal, referred to as the “off-line simulation” view (so-called because the simulation is not getting its inputs directly from the things being simulated), argues that we do not need to assume that people have a tacit theory of how other people’s minds work in order to anticipate what they will do or how they will feel under various situations. Instead all we need is to put ourselves in their position and ask what *we* would do. This way of putting it still leaves open the question of whether the latter predictions come from a special behavior-generating mechanism in our brain (our cognitive architecture) or from a tacit theory, and the difference is hotly debated in the case of social cognition (some of the arguments are reviewed in Nichols, Stich, Leslie, & Klein, 1996). These two alternatives differ only if the mechanism being appealed to is different from our general reasoning capacity. This in turn means that to be a true alternative to what some have called a theory-theory explanation, the simulation explanation must appeal to an encapsulated or modular architectural system. If it requires general reasoning, which is able to access any relevant knowledge, then the two options are indistinguishable.

Granted that the “simulation mode” of reasoning is used in many kinds of problem solving, two questions still remain: (1) Why should this mode be used at all, as opposed to some more direct way of solving the problem, and (2) When it is used, what does the real work of solving the problem – a special part of our mental architecture that deals with images or that generates behavior (or at least behavior plans), given a particular imagined situation – or inferences from tacit knowledge? I have already suggested one major reason why subjects might use the simulation mode in imagery studies: The task of imagining something invites it, since the task of “imagining X” is properly understood as the task of pretending that you are looking at situation X unfolding and you report the sequence of events that you would see (in the right order and in roughly correct relative times). Even without instructions from an experimenter, the simulation mode is often natural because of the nature of the task – e.g., if it is a task that you would normally carry out by making a series of actions and observing what happens, you might be tempted to imagine doing the same thing. Imagery is most often used for tasks that ask what would happen in a certain counterfactual (what if...?) situation involving perceivable spatio-temporal events.

In what follows I will sketch a number of extremely influential experimental results and compare explanations given in terms of inherent properties of the image (the architecture of the image system) and those given in terms of the simulation-from-tacit-knowledge explanation (and other considerations as well).

6.3.2.1 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 image scanning result is the following: it takes longer to “see” a feature in a mental image the further away it is from a place where one has been focusing. So for example, if you are asked to imagine a dog and inspect its nose and then to look at its tail, it will take you longer than if you were asked to first inspect its hind legs and then to look at its tail. Here is an actual experiment, perhaps the most cited result in all the

imagery research literature (first reported by Kosslyn, Ball, & Reiser, 1978). Subjects were asked to memorize a map such as the one in Figure 6-4. They are then asked to imagine the map and to focus their attention on one landmark on it – say the “church”. In a typical experiment (there are many variants of this basic study) the experimenter says the name of a second thing in the image (say, “beach” or “tree”), whereupon subjects must examine their image and to press a button as soon as they can “see” the second named place in their image. What Kosslyn (and many others since) found is that the further away the second place is from the place on which subjects are initially focused, the longer it takes to “see” the second place in their “mind’s eye”. From this result most researchers have concluded that *greater distances on the imagined map are represented by greater distances in some (mental) space*. In other words, they concluded that mental images *have* spatial properties – i.e., they have spatial magnitudes or distances. This is a strong conclusion about cognitive architecture. It says, in effect, that the symbolic code idea I discussed earlier 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; for example by saying the places are *n* meters (or units) from one another. But the representation of larger distances is not itself in any sense larger. The question then is: Is this conclusion about architecture warranted? Does the difference in time in this case reveal a property of the architecture or a property of what is represented? This exactly parallels the situation in the color-mixing example I discussed earlier where I asked whether a particular regularity 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 we need to determine whether the pattern of increasing scanning time arises from a fixed capacity of the image-representation or image-processing system, or whether the time it takes can be changed by changing the beliefs that subjects hold about how things are in the world or about what they are supposed to do in this experiment. In other words, we need to ask whether the empirical regularity is cognitively penetrable.

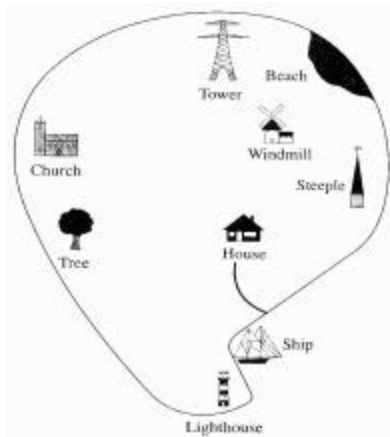


Figure 6-4: Example of a map to be learned and then imagined to study mental scanning

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 above map. Imagine that a light goes on at, say, the beach. Now imagine that this light goes off and instantly another light comes on at the lighthouse. Did you need to scan your attention across the image to see this happen – 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 mounted on a display board, with lights at the target locations, 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 the map and to indicate (by pressing a button) when they could “see” the second 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 time increase with increasing distance; subjects were not asked to imagine scanning that distance! 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 please since it is *your* image and *your* imagining.⁶ At least you can do these things to the extent that you know what it would be like if you were to see them and so long as you are able to generate the relevant measurements, such as the time you estimate it would take to get from point to point.

My proposal regarding the scanning experiment and many other such experiments involving mental images, is that the task of imagining invites observers to pretend (in whatever way they can) that they are looking at some situation and then to use whatever knowledge they have and whatever problem-solving technique seem to them most relevant, to generate the appropriate sequence of mental states (notice I do not say that they generate the answer they believe is wanted, but that they generate some sequence of mental states). The criticism of the scanning studies is not that the investigators have failed to control for something and have allowed an artifact to contaminate the results. It is that a proper understanding of the task requires that subjects try to do certain things; that is what is meant by a “task demand”. Other investigators have confirmed the relevance of task demands (Intons-Peterson, 1983; Intons-Peterson & White, 1981; Mitchell & Richman, 1980; Reed et al., 1983; Richman et al., 1979). Yet when these criticisms are discussed in the literature, they are often understood as the criticism that the results are due to the experimenter leading the subject, or that subjects were complying with the experimenter’s expectations. While this may often be true, the deeper point is that subjects are mostly just doing the task, as they understand it – i.e., pretending that they are looking at a map and seeing places on it (or seeing some sequence of events occur).

Notice that whether or not you choose to simulate a certain temporal pattern of events in the course of answering a question may 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 to be so obviously relevant to solving a particular problem. For example, we ran the following experiment that also required retrieving information from an imagined map by focusing attention on various locations on the map (Pylyshyn, 1981). Subjects were asked to memorize the same map as I described above and to refer to their image of the map in solving the problem. Rather than asking them to imagine looking at one place on the map and then to look for a second named place and indicate when they could “see” it (as in the original studies by Kosslyn et al., 1978), the task was instead to indicate the direction (in terms of a clock face) from the second named place to the previously focused place (for example, using the map in Figure 6-4, you might be asked to focus on the tower and then to indicate what direction the tower would be from the tree; the correct answer is “at 2 o’clock”). This direction-judgment task requires that the subject make a judgment from the perspective of the second place, so if anything requires focusing at the second place on the map this was certainly a good candidate. Yet in this experiment, the question of how you get from the first place to the second place on the map was less prominent than it was in the task was to

⁶ When we first carried out these studies we were (quite rightly, in my view) criticized 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 did (including the one described below) only demonstrate the obvious. That being the case one might well wonder what the fuss is 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.

“examine the map and indicate when you can see X”, especially when that request was presented right after the subject had been asked to focus on another place on the map. The present task required that subjects concentrate on the relative directions between two places once both places had been retrieved, rather than on how they got from one place to the next or on how far away the two places were on the map. In this case we found that the distance between places had no effect on the time taken to answer the question. Thus it seems that the effect of distance on reaction time is cognitively penetrable.

People have sometimes suggested that one can accommodate the finding that the “scanning effect” is cognitively penetrable by noting that the observed behavior depends on both the form of the image and the particular processes that use it, so that the differences in the process in this case might account for the different result one gets in different contexts. This is true; but then what explanatory work is being done by the alleged imagery medium and its alleged spatial property? Whenever we appeal to the nature of the process, then the hypothesized properties of the image plays no role. The claim has to be that under specified conditions the metrical properties of images is exhibited. But the specified conditions mustn’t be defined circularly. Saying that we only get the scanning effect when subjects imagine that they are scanning their attention from place to place, and not when they are hopping from place to place may well be true, but what about when they are imagining scanning fast or scanning slowly, or accelerating/decelerating as they scan? Is there anything that the image restricts them from imagining? The answer is surely that people can do whatever they wish. What, then, does claiming that images are spatial commit one to? So long as the assumption that there is a spatial representation that “preserves metrical distance” does not constrain possible behavior, it plays no role in the explanation, it remains theoretically irrelevant.⁷

Note only can observers move their attention from one imagined object to another without scanning continuously through the space between them, but we have reason to believe that they actually *cannot* move their attention continuously, as though smoothly tracking an imagined movement. In interviewing subjects after the study, many said that they did not feel they moved or scanned their attention *smoothly* through intermediate locations. They claimed that they did not notice (or “see”) objects at intermediate locations in their mental image in the course of their scanning, unless the objects were relevant to the task. This led us to wonder whether it was even possible to continuously scan an image as required by Kosslyn’s display-scanning model. The model claims that in getting from A to B on a mental image you must trace a path that takes you through all intermediate points (within the resolution of the image), just as a moving spot or a smooth eye movement would visit such places in the corresponding case of a real map. Without this assumption there would be no reason to predict that the time it takes to scan a certain distance should be proportional to the distance. But both introspection and objective studies suggest that we are unable to actually *move* an imagined point through a smooth set of invisible places along a mental scan path. In fact Jonathan Cohen and I (Pylyshyn & Cohen, 1999) carried out a number of experiments which show that observers are poor at judging where an imagined moving object is located at various times during its imagined passage through a region where there are no visible elements (e.g., in the dark). This suggests that the imagined object does not actually *pass through* a sequence of locations, a view defended later in this chapter (section 6.5.3).

Given that the times for mental scanning are actually computed, as opposed to observed in the mental image, one might well wonder why people bother to imagine a point of focus being scanned (even if not continuously) across their image? In the case of scanning while looking at a scene, I have already suggested that this might pay off because it allows people to use their superbly accurate capacity to compute “time to collision” based on a given speed of movement and a give (perceived) collision location. But it has been

⁷ Kosslyn insists that the medium does provide constraints. According to (Kosslyn, 1994, p11), “The subjects apparently can control some aspects of processing, such as speed of scanning, but not others; in particular, they could not eliminate the effects of distance on the time to shift attention across an imaged object.” But this is flatly contradicted by our data, as well as one’s personal experience of being able to hop around on one’s image. In fact, Kosslyn’s earlier model (Kosslyn, Pinker, Smith, & Shwartz, 1979) had a jump operation specifically to acknowledge that flexibility.

reported that people appear to use the scanning strategy even when not observing a scene (e.g., with their eyes closed). Why should they persist on using this method when scanning entirely in their imagination, where the time-to-collision computation may not be available (although it may actually be available from memory – people can compute time-to-collision after a moving objects, such as a baseball, disappears behind an occluding surface)? Assuming that people actually do use the mental scanning simulation with their eyes closed, a possible reason, already alluded to earlier, is that when using the simulation mode to imagine solving a problem, people may carry over strategies that they use in real situations, where the strategy is clearly the appropriate one. Inasmuch as they are being asked to imagine that they are seeing something, it makes sense to imagine what they would have done in the visual case.

Consider the example presented by (Pinker, Choate, & Finke, 1984), in which people claim they “extrapolate” an arrow, whose location and direction is not seen but retrieved from memory, in order to determine whether the arrow points at a particular dot in a recalled image. In this task Pinker et al. found an effect of distance on response time (although interpreting this result is not unproblematic inasmuch as the response times involved are very much longer than reported mental scanning times, see Pylyshyn, 2002). But extrapolating a line from the arrow is clearly the right strategy *if you have ruler and a rigid surface* since in that case you would be able to see whether the extrapolated line would intersect the dot. It may also be the right strategy to use if you can see the arrow and the dots, since that judgment can make use of some of our precise psychophysical skills, such as judging whether several visual elements are collinear or whether two line segments are aligned, as are in the case of vernier acuity measurement (as with the time-to-collision task, it is not known how this is done in detail, only that people are fast and accurate at it). It may even be the right strategy to use when a perceived arrow is superimposed on a perceived screen and subjects have to judge whether the arrow points to dots that had been on the screen shortly before (as was the case in the earlier experiment by Finke & Pinker, 1982), since in that situation subjects can use their ability to visually index features in the display (even a uniform screen), their high capacity visual short term memory, together with their ability to compute time-to-collision. But in the purely imagined case (where both arrow and the dot must be recalled from memory) the problem for the theorist is not to explain why that strategy was used (it was probably used because it would have been the appropriate thing to do in the situation being imagined and the subject may be simulating this strategy), but rather to explain how such a strategy could possibly work without a real display. Unless you believe in a literal picture theory (in which case you have a lot of other problems, as we will see in the next chapter), drawing a mental line in mental space cannot solve the problem because mental lines do not actually *have* the property that when they are extrapolated they intersect appropriately situated mental objects. In imagining the line being drawn you don’t actually draw anything, all you do is *imagine that* (i.e. think the thought that) the line is initially short and gets progressively longer. And you don’t need a special form of representation to do that. All you need is a representational system capable of representing (not of *having*) various lengths and distances. But of course so long as there is no physical line on a physical surface, no perceptual judgments can be based on such thoughts (which may account for the strong allure of the literal picture theory). Therefore the judgments are being made by some other means, which, as usual in such cases, is not available to introspection. This question is closely relate to the general question of where the spatial character of mental images comes from and I will take up the question in the next chapter (especially in section 7.3.2).

6.3.2.2 The “size” of mental images

Here is another example of a widely cited result that might well be attributed to the use of tacit knowledge to simulate the imagined event. Consider the finding 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, see Kosslyn, 1980, chapter 9). This seems a clear candidate for being the result of an implied requirement of the imagery task. For if you are asked to imagine something small then 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 involves. One reason you might do this is that when you actually *see* a small object (or an object that is far away) you can make out fewer of its details due to the limited resolution of your eye. If you are to accurately simulate the visual experience of seeing an object, then in the case of a small object you have to take additional steps to make the details available (e.g., imagine bringing the object closer or zooming in on it). But what does it mean to make your image “larger”? Such a notion is obviously meaningful only if the image *has* a real size or scale. If, as in our null hypothesis, it is like a description, then size has no literal meaning. You can *think of* something as larger or smaller, but that does not make some *thing* larger or smaller. 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 working memory or your active representation. So if this is what the task demands when the mouse is imagined as “large” then 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 counteract the effect of the “image size” manipulation, making details easier to report in small images than in large ones and vice versa? Can you imagine a small but high resolution or very detailed view of an object, in contrast to a large but low-resolution view or a view that for some reason lacks details? Surely you can, though I know of no one who has bothered to carry out an experiment such as asking subjects to report details from a large blurry image versus a small clear one. There is a good reason to forgo such an experiment. Consider what it would mean if such an experiment were done and showed that it takes longer to report details in a large blurry object than in a small clear one (as I am suggesting it would)? Would this show that it is the presence of visible details rather than size that is the relevant determiner of response time? Surely we expect that examining a blurry object will lead to difficulty in reporting its fine-grained properties. To see that there is a semantic issue involved here, think about what it would mean if, instead, people were *faster* in reporting details from a “blurred” mental image than a clear one or if it was faster to report details from a small image than from a large image. 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 small versus a large mental image, or a blurred versus a clear image. Such results would be incompatible with what happens in *seeing*. If one failed to see fine details in a 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 or on a noisy TV, and so on. As long as examining a visual image *means* simulating what it is like to see something, this must necessarily be the case.

Given the rather obvious reason for the parallels between vision and imagery in such cases one might even wonder how studies of mental image inspection could fail to show that it parallels the case of seeing, assuming that observers know more-or-less what it would be like to see the object that is being imagined? This would apply to any property of seeing of which observers have some tacit knowledge or recollection. For example, it applies to the finding that the acuity map of mental images appears to roughly duplicate the acuity map of vision (Finke & Kosslyn, 1980). As noted earlier, 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 be before they cease to be discriminable, and it is not surprising that this is duplicated in imagery, especially when subjects are asked to turn their head (with eyes closed) and pretend to be looking at objects in their visual periphery. This methodology also illustrates the well-known phenomenon that recall is

better when the recollection takes place in an environment similar to the one that obtained in the situation being recalled.



Figure 6-5. The point of this cartoon is self-explanatory: We must not confuse the content of representations with their intrinsic property. (S. Harris, from *American Scientist*, 66, 647; reprinted with the artist's permission).

6.3.2.3 Mental “paper folding”

The examples discussed so far suggest that many of the mental imagery results may be due to subjects' simulating what they think would happen if they were witnessing the imagined event taking place. But why should people go through the trouble of simulating a situation if they already know (albeit tacitly) what the answer is? Several reasons were considered earlier (section 6.3.2) including the implied task demands. But there is another important reason we have not yet discussed, related to how the relevant tacit knowledge is organized. Consider the question: What is the fourth (or n 'th) letter after “M” in the alphabet? To answer that question, people normally have to go through the alphabetical sequence (and it takes them longer the larger the value of n). (This works even if the question is “Is R before M in the alphabet?” – the time it takes depends on how far apart the two letters are in the alphabet).

A somewhat more complex example, which may also well involve the same principal, applies to a paper-folding task studied by (Shepard & Feng, 1972). In their experiment, subjects are asked to mentally fold pieces of paper shown in a drawing (examples of which is shown in Figure 6-6), and to report whether the arrows marked on the paper would touch one another. Try this yourself. You will find, as they did, that the more folds it would require to actually fold the paper to see whether the arrows coincide, the longer it takes. Shepard & Feng took this to indicate that working with images parallels working with real objects. Elsewhere

(Shepard & Chipman, 1970) call this principle “second order isomorphism” and claim that it is a general property of mental images (though I have argued that such a principle is true of any empirically adequate functional model, regardless of the form of representation used, see Pylyshyn, 1984a, p 203).

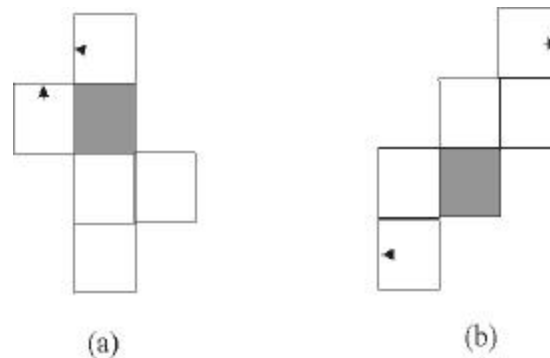


Figure 6-6: Two of the figures used in the (Shepard & Feng, 1972) experiment. The task is to imagine folding the paper (using the dark shaded square as the base) and say whether the arrows in these two figures coincide. The time it takes increases with the number of folds required.

The question we need to ask about this task is the same as the question we asked in the case of 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 to fold the corresponding real piece of 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. Possessing the relevant tacit knowledge would explain why it was possible for subjects to imagine folding the paper at all. As in the earlier examples, if you are asked to imagine folding a sheet of paper then you are required (on pain of failing to follow the instructions) to imagine each individual step that the folding goes through. But in the present case one would presumably get the same result even if one did not ask the subject to imagine folding the paper. Yet it is still hard to see how you could answer this question without imagining going through the sequence of folds. Why should this be so? A plausible explanation, which does not appeal to properties of a special imagery system, is that the observed time difference has to do with how knowledge of the effect of folding is organized – just as was the case in the alphabet example. What we know by rote 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 (in the real world) would require 4 folds, we have to apply our one-fold-at-a-time knowledge four times. Recall the parallel case with letters: In order to say what the fourth letter after M is we have to apply the “next letter” rote knowledge four times. In both the alphabet case and the paper folding case subjects could presumably memorize the results of macro-operations. They could commit to memory such facts as which letter of the alphabet occurred two (or n) letters after a given letter, or in the case of paper folding task they could memorize such facts as what results from double folds of different types. If that were how knowledge was organized, the results discussed above would most likely no longer hold. The important point is that, once again, the experimental result may tell us what sequence people go through in solving a problem (that’s the “second order isomorphism” phenomenon) but it tells us nothing about *why* people must go through that sequence rather than some other. It also does not tell us about how the states of the problem are represented and whether one needs to appeal to any special properties of image representations to explain the results. In this case it tells 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. Knowledge 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 an image of that person. 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).

6.3.2.4 Mental Rotation

One of the earliest and most cited results in the research on the manipulation of 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 6-7, and asked them to judge whether the two objects depicted in the drawings were identical, except for orientation. Half the cases were mirror-images 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 (except when the angle was over 180 degrees, when in some cases the time it took became proportional to the angular distance remaining – i.e., to the angle measured counterclockwise).

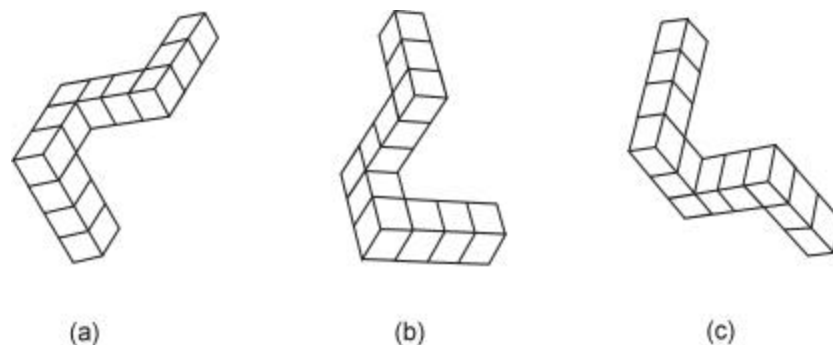


Figure 6-7. Examples similar to those used by (Shepard & Metzler, 1971) to show "mental rotation." The time it takes to decide whether two figures are identical except for rotation, as in the pair (a-b) or are 3D mirror images (enantiomorphs), as in the pair (a-c), increases linearly as the angle between them increases.

This result has been universally interpreted as showing that images 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 what it means to say that the image is rotated – what exactly is rotated – and 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. For example it might be the case that during mental rotation the figure moves *as a rigid form* through a continuum of angles, thus capitalizing in an intrinsic property of the image format that maintains the form in a rigid manner. Also relevant to this explanation is the question of whether we are *required* to apply this particular transformation in order to compare two shapes in different orientations. Not surprisingly, it turns out to be important that the mismatches be ones in which the figures have the same features but are enantiomorphic images, otherwise observers simply look for distinguishing features and no mental rotation ensues (Hochberg & Gellman, 1977). There are two points to be made about this mental rotation phenomenon.

First, contrary to the general assumption, the figural “rotation” could not be a holistic process that operates on an entire figure, changing its orientation continuously while rigidly retaining its shape. In the original 3D rotation study (Shepard & Metzler, 1971), the two comparison figures were displayed at the same time. A record of eye movements made while doing this task reveals that observers look back and forth many times between the two figures, checking for distinct features (Just & Carpenter, 1976). 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 present, no “rotation” is found. In studies reported in (Pylyshyn, 1979b) I showed that what counts as the “rate of rotation” (the rate of change of reaction time as a function of the angle of mismatch, as seen in the slope of the graph plotting the time versus the mismatch angle) 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 or not a misoriented test figure was embedded in the original figure, as shown in Figure 6-8). The fact that the apparent “rate of rotation” depends on such organizational and task factors shows that whatever is going on in this case does not appear to consist in merely “rotating” a shape in a rigid manner until it is in rough correspondence with the reference figure.

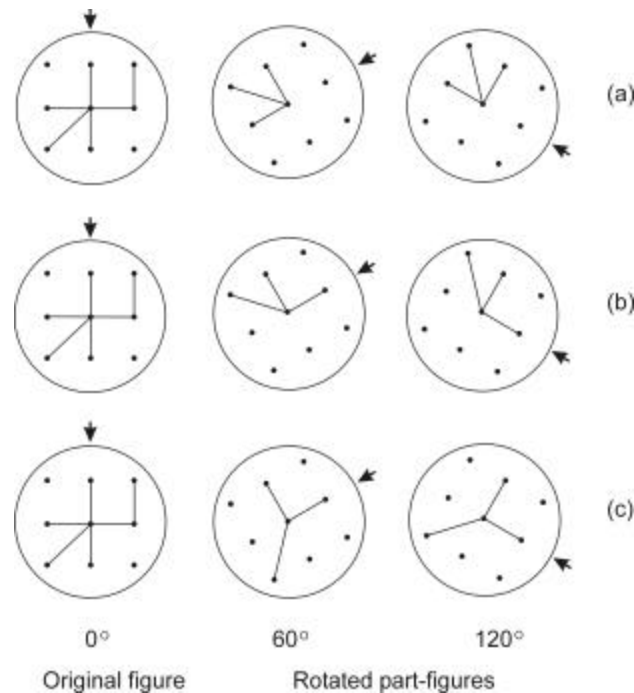


Figure 6-8. Figures used in (Pylyshyn, 1979b) to show that mental rotation is not a holistic process. The task was to indicate whether the figures in the second and third columns were parts of the figure in the first column. The part-figures varied in how “good” a subpart they were. Row (a) shows a “good” subpart, row (b) shows a “poor” subpart and row (c) shows a figure that is not a subpart of the original figure. Results showed a “rotation” effect (larger angles took longer) but the “rate of rotation” depended on how good a subpart the probe part-figure was.

Second, 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 notion that a representation maintains its shape because of the inherent rigidity of the image while it is rotated cannot literally be what happens, notwithstanding the phenomenology. Since the *representation* is not literally being rotated – neither the brain cells that encode the figure nor any other form of encoding is being moved in a circular motion – the closest to a “rotation” that might 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. There are undoubtedly good reasons, based on computational resource considerations, why the process might proceed by iterating over successive small angles (thus causing the comparison time to increase with the angular disparity between the figures) rather than attempt the comparison in one step. For example, small rotations, at least in 3D, result in small relative displacements of component parts of the form and decrease the likelihood of a new aspect coming into view.⁸ As a result incremental rotation might require less working memory to ensure the maintenance of the relative location and connectivity of parts of the figure, using a constraint propagation method such as

⁸ An *aspect* is a topological structure of the edges, junctions, cusps, and other discontinuities, of a single viewpoint of an object. As the point of view changes a very small amount an aspect is likely to remain the same – constituting a set of stable vantage points. With larger changes new discontinuities come into view resulting in sudden changes in aspect, called *events*. The graph of discrete aspects and their interconnections is termed a graph of the *visual potential* (sometimes called an aspect graph) of an object and has been used in computer vision for characterizing a shape by its potential aspects (Koenderink, 1990). Going from one aspect to a very different one in the course of recognizing a complex object is computationally costly, consequently in recognizing figures from two different vantage points it would make sense to move gradually through the aspect graph and traversing adjacent aspects one at a time. This suggests that one should compare pairs of misaligned shapes by a gradual aspect-by-aspect change in point of view, resulting in a discrete sequence that may appear as a “rotational”.

commonly used in computer vision systems (and discussed in section 3.1.1.1). Recognition of constraints on working memory led Marr and Nishihara to hypothesize what they called a SPASAR mechanism for rotating simple vertices formed by pairs of lines and obtaining their projection onto a new reference frame (see Marr & Nishihara, 1976; 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 entailed a limited analogue operation on a small feature of a representation. 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 particular operation on parts of a structured representation (which might even be viewed as an analogue operation) that was responsive to a computational complexity issue.

In the context in which two differently oriented figures are being compared for whether they are identical or mirror images, the linear relation between angle and time seems to be robust and does not appear to be cognitively penetrable. It is thus not a candidate for a straightforward tacit knowledge explanation (as I tried to make clear in, Pylyshyn, 1979b). Rather, the most likely explanation is one that appeals to the computational requirements of the task and to general architectural (e.g., working memory) constraints. It therefore applies regardless of the form of the representation. Nothing follows from the increase in time with angle concerning either the form of the representation of the figures or of inherent properties of a representational medium, which assures the maintenance of the shape of an object through different orientations. Nor does the empirical finding suggest that the representation must be manipulated as a whole and must pass smoothly through intermediate angles. The problem of accounting for the linear relation of the angle between figures and the time to make the comparison is not resolved by talk or image “rotation.” After all, we still need to say what, if anything, is rotated in mental rotation experiments. The only way that an image rotation” story can be explanatory is if we are willing to attribute causality to the phenomenal experience, which few people are willing to do once it is pointed out that this is what the account entails.

6.3.3 A note concerning *cognitive penetrability* and the appeal to tacit knowledge

How can you tell whether certain patterns of observations made while a subject is solving a problem tell us about the nature of the architecture of the imagery system or about the person’s tacit knowledge and the way it is organized? One diagnostic I have advocated, and discussed at length in (Pylyshyn, 1984a), is to test for the cognitive penetrability of the observed pattern. Because many of my criticism of particular conclusions that have been drawn from imagery research have been focused on this criterion, this criterion has been criticized by picture-theorists. And because so much of this criticism has been based on misunderstandings of the role that this criterion plays, I will devote some space to a discussion of this issue.

If the reason for a particular pattern of observations (say of the pattern of reaction times for different variants of the task) is that people are simulating a situation based on their tacit knowledge of what it would look like, then if we alter the knowledge or the assumptions about the task, say by varying the instructions so as to change people’s beliefs about the task situation, the observations may change accordingly, in a way that is rationally connected with the new beliefs. This is the basis of the criterion of cognitive penetrability. For example, if we instruct a person on the principles of color mixing we would expect the answer to the imaginal color-mixing question discussed earlier to change appropriately. Similarly, if you ask people to tell you when they see certain details on an image after they have been focusing on a certain place, then depending on subjects’ beliefs, the time it takes may be altered. For example, if subjects do not believe that they have to scan their attention to the see the details, or if they believe that in the visual situation a light will come on immediately after the second place is named, then there may not be a linear relation between distance on the image and reaction time (as we showed experimentally; see section 6.3.2.1). But as is the case with all

methodological criteria, cognitive penetrability cannot be applied blindly. Even when some behavior depends on tacit knowledge, the behavior cannot always be appropriately altered with instructions, and you certainly can't discover what tacit knowledge subjects have by merely asking them. This is a point that has been made forcibly in connection with tacit knowledge of grammar or of social conventions, which also typically cannot be articulated by members of a linguistic or social group, even though violations are easily detected.

In using the cognitive penetrability criterion, it is important to recognize that the fact that an imagery-related pattern of observation is cognitively impenetrable does not mean that the pattern arises from properties of the mental image or of image-manipulating mechanisms. Many behaviors that are clearly dependent on specific knowledge are immune from cognitive influence (e.g., phobias and obsessive behaviors). In general, when people are shown that some belief they hold is false they do not immediately change it; beliefs are generally resistant to change so being impenetrability by relevant information is not sufficient for concluding that some regularity does not arise from tacit knowledge but rather from a property of the architecture of the image system. Moreover, as I noted earlier, not every aspect of patterns observed in mental imagery experiments are due to tacit knowledge and therefore not all aspects of these patterns need be cognitively penetrable. Consider the following example in which the cognitive impenetrability of some aspect of a pattern of observation has been taken as evidence against the tacit knowledge explanation of the entire phenomenon. In a series of experiments on mental scanning, (Finke & Pinker, 1982) found that the time required to judge that an arrow points to a memory image of a dot increases the further away the dot is from the arrow. Nonetheless, Finke et al. argued that this case of mental scanning could not have been due to tacit knowledge (contra my claim, discussed in section 6.3.2.1). The reason they give is that although 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. Of course neither could the authors – and the reason is that the aberrant short-distance time was most likely due to some entirely different mechanism (perhaps the effect of “attentional crowding” in the original visual display) from that which caused the monotonic increase of time with distance, a mechanism not known to either the subjects or the experimenters. A major reason why one sometimes observes phenomena in imagery studies that are not attributable to knowledge is precisely the reason recognized by (Finke & Freyd, 1989): most observed phenomena have more than one cause. Even when tacit knowledge is the main determiner of the observed pattern, other factors also contribute. Because of this, cognitive impenetrability has to be used to examine specific proposed mechanisms, not entire experimental results. Moreover, cognitive penetrability is a *sufficient but not necessary condition* for attributing a pattern to the tacit knowledge held by subjects in these experiments. If a particular pattern is penetrable by changing beliefs we can conclude that it is due to tacit knowledge but if it is not penetrable it may or may not be due to tacit knowledge and it may or may not be attributable to some property of the imagery architecture.

Another example of the mistaken inference from cognitive impenetrability to assumed properties of the imagery system concerns 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 system (i.e., to the nature of the imagery architecture). However, (Ranney, 1989) suggested that the phenomenon may actually be due to tacit knowledge. Yet it seems that at least some aspects of the phenomenon is not cognitively penetrable (Finke & Freyd, 1989). Does this mean that representational momentum must then be a property of the imagery architecture, as Freyd & Finke assumed? What needs to be recognized, in this and other such cases, is that there may be many other alternative explanations besides tacit knowledge or imagery architecture. In this particular case there is good reason to think that part of the

phenomenon is actually visual and takes place during perception of the moving object, rather than in its memory representation. There is evidence that the perception of the location of moving objects precedes the actual location of such objects (Nijhawan, 1994). Eye movement studies also show that gaze precedes the current location of moving objects in an anticipatory fashion (Kowler, 1989, 1990). Thus even though the general phenomenon in question (the forms of visualized motion) may be attributable to tacit knowledge, the fact that in these studies the moving stimuli are presented visually may result in the phenomena also being modulated by the visual system operating on the perceived scene. The general point in both these examples is that even genuine cases of failure of cognitive penetrability do not mean that the phenomena in question reveal properties of the architecture of the imagery system.

A final point that needs emphasizing is that no criterion can be applied blindly to the interpretation of empirical results without making collateral assumptions. No cognitive mechanism (nor, for that manner, no physical law) can be observed directly. Even something as clearly a part of the architecture as a “red” detector, may or may not be deployed in a particular situation, depending on the observer’s beliefs and goals. A red detector can be used to find a red car in a parking lot, yet the overall task of looking for the car is clearly cognitively penetrable (it matters what color you believed the car was and where you believe you left it). To distinguish the functioning of the detector from the functioning of the cognitive system in which it is embedded, we need to set up the appropriate control conditions and we also need to make some (independently motivated) assumptions about how the task is carried out. As I have repeatedly said (e.g., Pylyshyn, 1978) observed behavior is a function of the *representation-process* pair, and one cannot be observed without the other. But that does not mean (as Anderson, 1978, has concluded) that we cannot in principle decide the nature of either the process or the structure of the representation. As I argued in my reply to Anderson (Pylyshyn, 1979c), the situation here is no different from the one that applies in any science, where theory is always underconstrained by data. The history of information-processing psychology (as well as in other sciences) has shown that the problem of deciding between alternative theories is not principled, but only practical (at least outside of quantum mechanics). Cognitive penetrability is not a panacea, but a tool, much like reaction time or fMRI, to be used in carefully controlled experiments. While one can always argue about individual cases, the cases of cognitive penetration that I have cited here and elsewhere (Pylyshyn, 1981, 1984a) seem to me clear cases that show that particular hypotheses about the constraints imposed by the mental imagery mechanism are not tenable.

6.3.4 Summary of some possible reasons for observed patterns of imagery findings

As I mentioned throughout this section, there is more than one reason why particular systematic patterns of observations are found in studies of mental imagery, other than inherent properties of the imagery system (i.e., the architecture of mental imagery). Here is a summary of some of the reasons discussed in this section concerning why imagery experiments may produce the pattern of observations they do.

(1) First and foremost, the pattern may be due to the *use of tacit knowledge to simulate aspects of real-world events, as they would appear if we were to see them unfold*, including the stages through which such events would proceed and their relative durations. Subjects may cause the observed results to come out in certain ways because they know how things in the world work and because they interpret the task as that of simulating how things would look to them if they were actually to happen in the world. We call the body of facts and beliefs that are brought to bear in this simulation, *tacit knowledge* because people may not be able to articulate what they know – as is generally the case with knowledge of naïve physics, of the conventions of social interactions, of the structure of one’s language, and of what motivates and guides other people’s behavior. Although such knowledge must be inferred indirectly by observing its consequence on behavior it

must also meet the criteria of being knowledge – it must enter into inferential processes (so that its effect depends on what else observers know and what their utilities and goals are) and have generalized effects beyond a narrow domain.

(2) The pattern may be due not only to the content of tacit knowledge, but also to *how this knowledge is organized*. It is commonplace in psychology to find that in order to solve a problem people may be forced to go through a sequence of steps dictated by the way their knowledge is organized. A simple example of this is the task of saying what the n th letter is after a given letter in the alphabet. The time it takes to carry out this task depends on the fact that our knowledge of the alphabet happens to be organized in the form of a list to be accessed in sequence (presumably because it was learned that way). Similarly, the pattern we observe may arise from the fact that in using imagery certain access paths to relevant knowledge may be easier or more salient. It is also a commonplace finding in studies of reasoning that rewording a problem in a different but logically equivalent way may result in it being either much easier or much more difficult to solve (Hayes & Simon, 1976; Kotovsky, Hayes, & Simon, 1985). A prime candidate for this sort of explanation of an observed effect is the phenomenon concerning mental paper-folding, discussed in section 6.3.2.3.

(3) The pattern may be due to a *habitual way of solving certain kinds of problems* or to more frequent patterns of solution. This is the basis for the widely-studied phenomenon known in the problem-solving literature as “mechanization”, an example of which is the Luchins water jug experiment (Luchins, 1942). In this example subjects are shown a series of problems concerning measuring out a specified amount of liquid using containers of certain sizes. After successfully solving a series of problems using a certain combination of containers, subjects were unable to solve a new (equally simple) problem requiring a different combination of containers. Another is the so-called “functional fixedness” phenomenon studied by (Anderson & Johnson, 1966) in which people get stuck with one view of the function of an object (say viewing a matchbox as a container) and have a great difficulty solving a problem that is easily solved by viewing the same object in a different way (say as a potential shelf). This sort of effect is also common with tasks involving imagery, where we tend to use a particular, often habitual, way of imagining a situation – particularly when we have a habitual way of solving a problem in reality when we can actually carry out certain actions and watch the solution emerge (as in the case we discussed earlier of determining whether a line can be extrapolated to intersect a visible object).

(4) The pattern may be due to the nature of the task itself. As (Ullman, 1984) has argued, some tasks logically *require* a serial process for their solution. Depending on what basic operators we have at our disposal, it could turn out that the stages through which a solution process passes and the time it takes may be essentially determined by task requirements. (Newell & Simon, 1972) have also argued that what happens in certain segments of problem solving (as it shows up, for example, in what they call a “problem behavior graph”) reveals little about subjects’ strategies because in those episodes subjects’ choices are dictated primarily by the demands of the task – i.e., subjects are doing what must be done to solve the problem or the obvious thing that any rational person would do in that state of the problem solving process.

(5) The pattern may also be due to general computational complexity constraints which result in trading off increased time for less complex operations, as hypothesized for mental rotation by the SPASAR theory (Marr & Nishihara, 1976), or as hypothesized by (Tsotsos, 1988) in connection with the question of why certain operations are carried out sequentially rather than in parallel. A candidate for this sort of explanation is the mental rotation phenomenon discussed in section 6.3.2.4.

(6) It is also possible that when we carry out experiments on mental imagery we will find effects that are not

due solely to one of the above reasons, but to a combination of reasons. Phenomena may have multiple causes; it may be due to the interaction between the tacit knowledge used in the mental simulation and also certain properties of the cognitive architecture involved in reasoning. The architectural properties may be general ones (e.g., limitations of the short-term memory in which image-representations are stored during processing) or ones that are specific to creating and manipulating visual mental images. What we need to do is determine whether tacit knowledge will explain the effect since if it can we are left with the “null hypothesis”. If it cannot, then we need to look for other effects, including interactions.

Having laid out some of the issues, as well as some alternative explanations of certain well-known empirical observations associated with the study of mental imagery phenomena, I now turn to a more detailed examination of some of the theoretical views concerning what makes mental imagery a special form of cognitive activity – views that have now become as close as one can get to being received wisdom within psychology and parts of philosophy of mind.

6.4 Some alleged properties of images

6.4.1 Depiction and mandatory properties of representations

It has been frequently suggested that images differ from symbolic forms of representation (such as those sometimes referred to as “language of thought”) in that images stand in a special relationship to what they represent, a relationship often referred to as *depicting*. One way of putting this is to say that in order to depict some state of affairs, a 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,⁹ so I quote from him at some length (Kosslyn, 1994, pp5).

“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). Resemblance is neither necessary nor sufficient for something to have a particular reference: Images may resemble what they do not refer to (e.g.

⁹ 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 developed theory of mental imagery and has been 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 theory can be taken as the received view in much of the field.

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 such obvious problems, the notion of resemblance keeps surfacing in discussions of mental imagery, 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 in many ways 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 your image looks like, what conscious content it 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 (i.e., what form and substance underlies it), but surely you cannot be deceived about what your mental image looks like, or what it resembles. That is not an empirical fact about imagery, it's just a claim about what the phrase "mental image" means.

What gives representations a role in thought is the fact that if it is possible for processes in the organism to obtain relevant information from them, and to make inferences from them, thereby making explicit some otherwise implicit information. There might be a weak sense of "resemblance" wherein a representation can be used to provide information about the appearance of an object or to allow such information to be made explicit when it was only implicit (for more on the question of how *real* visible images allow one to draw certain kinds of inferences, see section 8.3, especially 8.3.6). But in this sense every representation of perceptible properties (including a description of how something looks) could be said to resembling its referent. Any stronger sense of resemblance inevitably imports an intelligent and knowledgeable agent for whom the representation "looks like" what it represents (i.e., what it would look like if perceived visually). This agent is the means by which one eliminates the inherently many-to-many relation between an image and its meaning that Wittgenstein and others talked about. As Wittgenstein reminded us, an image of a man walking upstairs is the same as the image of a man walking downstairs backwards. An intelligent agent could select one of these interpretations over the other, but would do so on the basis of reasoning from other (non-image) considerations.

In contrast to the problematic criterion of resemblance, the proposal that images are decomposed into "parts" with the relations among parts in some way reflecting the structure of the corresponding parts of the world, does deserve closer scrutiny. It is closely related to a criterion discussed by (Sloman, 1971), although he suggested this as characteristic of analogue representations. Fodor and I referred to this sort of part-whole structure as the compositional character of representations and claimed that it is a requirement on any form of representation adequate to explain the representational capacity of intelligent organisms and to explain the capacity for thought and inference (Fodor & Pylyshyn, 1988). Thus if images are to serve as vehicles of thought, they too must be compositional in this sense. And if they are compositional then they have what might be called interchangeable parts, much as lexical items in a calculus do. This however makes them more language-like (as in the "language of thought" proposal of Fodor, 1975) than pictorial and says little about the alleged depictive nature of images since it applies equally to any form of representation. Indeed, it is doubtful that images are compositional in the required sense, even if they do have parts.

Another proposal in the Kosslyn quotation above is that in depictive representations it is mandatory that

certain aspects be made explicit. For example (according to Kosslyn) 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 you about the form of the representation remains open. It is in fact a general property of representations that some aspects tend to be encoded (or at least assigned as a default) if other aspects are. Sometimes that is true by definition of that form of representation or by virtue of the logical entailments of certain facts that are represented. So, for example, you can't represent a token of a spoken or written sentence without making a commitment as to how many words it has, you can't have a representation containing four individual objects without implicitly representing the fact that there are at least four of them, and so on. Of course the converse is not true; the sentence "there are four plates on the table" does not contain four distinct representations of individual plates. Also the tendency to represent certain clusters of properties sometimes may be just a matter of habit or of convention or a reflection of the frequent co-occurrence of the properties in the world: When you represent someone as riding a bicycle you may also represent them as moving the pedals (even though you needn't have), when you represent someone as running you might also represent them as moving quickly, and so on, none of which is actually mandatory even if they are plausibly true. It may also be the case that certain patterns are frequent (if not mandatory) in visual images simply because they are frequent in the world being represented.

So the question is, when you represent some object in the form of an image is it mandatory that you represent its shape and size, and if so why? What about its color and 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? Is there something special about the encoding of shape, orientation and size of an object? We know that retinal size and retinal location can be factored away from the representation of an object (in fact it is hard to demonstrate that these are even encoded into long term memory), but can shape and orientation of an object also be factored away? Studies in rapid search suggest that we can identify the presence of a shape without identifying its location and we can identify both color and shape but miscombine them to form "conjunction illusions" (Treisman & Schmidt, 1982). In fact, these studies appear to show that in representing shape, abstract properties such as having a "closed contour" may be factored apart from other properties of shape and miscombined. While these studies do not tell us which properties must be contained in an imaginal representation, they do suggest that in the process of visual encoding, such properties as shape, color, location and even closure can be factored apart from one another (i.e., they are represented as separate codes). In Chapter 4 (and in Pylyshyn, 2001) I suggested that very early in the visual process, all such properties (shape, color, location) are factored from (and are initially secondary to) the individuality of visual objects. What Kosslyn may have had in mind in the earlier quotation is that when you ask someone to imagine an object, say the letter "B," the person will make a commitment to such things whether it is in upper or lower case. It does seem that you can't imagine a "B" without imagining either the upper case letter "**B**" or a lower case letter "**b**". But is this not another case of an implicit task requirement? Are you not being asked to describe what you see when you see a *printed token of a particular letter*? If you actually saw a printed token of a letter you would have to 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 what would you conclude? You might be tempted to say that the person did not have a visual image at all, but only some idea of the letter.

Yet most of our representations are noncommittal in many different ways (see Chapter 1 for examples).

In particular they can be noncommittal in ways that no picture can be noncommittal. Shall we then not call them images? Is an image generated in response to the letter-imaging example mentioned above not an image if it is non-committal with respect to its color or font or whether it is bold or italic? Such questions show the futility of assuming that mental images are like pictures. As the graphic artist M.C. Escher once put it (Escher, 1960, p7), "...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'."

One of the most important claims of the Kosslyn proposal, as expressed in the above quotation, is the idea that although images are inherently spatial, the space in question need not be physical but may be "functional". Both parts of this proposal (that images are spatial and that the relevant space might be a functional one) have been extremely influential and have led to new lines of research, some of which has involved neural imaging. The claim that images are spatial and the new lines of research that focus on this claim, will be discussed in detail in Chapter 7. For the present I will take up the question of whether there is any evidence to support the claim that images are the sorts of things that can be examined visually, as is clearly implied by the notion of "depiction" as a mode of representation.

6.5 Mental imagery and visual perception

Perhaps the most actively pursued question 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 for a number of reasons, not the least of which is the fact that the experience of mental imagery is 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 can be indistinguishable (Perky, 1910). I will return to the similarity of the experience of vision and imagery later (section 6.5.5) when I raise the question of what significance ought to be attached to this experiential evidence. In this section I look at some of the psychophysical evidence for the involvement of vision in imagery. In Chapter 7 I will consider a new, and to many investigators a much more persuasive class of evidence for the involvement of vision in mental imagery, evidence that derives from neuroscience (especially neuroimaging studies and clinical reports of brain damaged patients).

In examining the question of whether (and in what way) vision may be involved in mental imagery, it is important to make a clear distinction between the visual system and the cognitive system, since cognition clearly is involved in some stage of both mental imagery and what we generally call visual perception. In order to see how intimately imagery involves the visual system we must first provide some criteria for when we are observing the operation of the visual system, as opposed to the system that includes reasoning and cognition in general. This is why we need to have some characterization of the proprietary properties of the encapsulated or cognitively impenetrable aspects of vision. We need to identify a narrower technical sense of "visual system," as I attempted to do in chapters 2 and 3 (and in Pylyshyn, 1999) where I used the term "early vision" (a term probably introduced by David Marr) to designate the proprietary modular part of vision that is unique to that modality. In investigating the involvement of the visual system in mental imagery we must also distinguish between effects attributable to the operation of a special visual architecture from effects attributable to the fact that in visual imagery we are concerned with different subject matter, for example we are typically concerned with visual (optical, geometrical) properties of some scene (actual or hypothetical). The "null hypothesis" strategy I introduced earlier (section 6.1) (i.e., that in the absence of evidence to the contrary we shall assume that all thoughts take the same symbolic format) says that we need to ask whether a

system that did not have a special *form* of encoding or a special architecture, might nonetheless exhibit the observed characteristics when it reasoned about visual properties. We might, after all, expect that certain distinct characteristics might be exhibited when we reason about some special subject matter, such as about emotions, feelings, interpersonal relations, or mathematics, or perhaps when we reason about such psychologically distinct categories as animate versus inanimate things (e.g., there is reason to think that such subject matters may even involve different parts of the brain, see Dehaene, 1995; Samson, Pillon, & De Wilde, 1998). What this means is that we have to insist that certain distinctions be honored when asking whether imagery involves the visual system. In what follows I will examine a number of lines of evidence that have persuaded people that mental imagery involves the visual system, while keeping in mind the need for a finer technical distinction between “visual system” and the rest of the cognitive mind.

Many of the experiments described earlier (including the image scanning experiments and the studies involving projecting images onto visual stimuli) have been interpreted as suggesting that mental images are inspected by the visual system. I have already discussed the finding that it takes longer to judge the presence of a small feature in a small image than in a large one. There is also a well-documented relation between the relative size of a pair of imagined objects and the time it takes to judge which one is larger. For example, when we make certain judgments by examining a visual image we find the same psychometric function as we get when we do the task visually: it takes longer to judge (from an image) whether a toaster is larger than a person’s head than to judge whether a toaster is larger than a horse. The fact that this relation is the same as it is when real objects are being viewed suggested to some people that the same visual mechanisms were being used in both cases. Although this phenomenon received a great deal of attention in early work on mental imagery, it soon became clear that it had nothing to do with the putative *visual* aspect of mental images since the effect occurs with any comparison of magnitudes, including judgments of such abstract properties as the relative cost or attractiveness of different objects, or the relative magnitude of numbers (it is faster to judge that 374 is larger than 12 than that 21 is larger than 19). This more general phenomenon is called the “symbolic distance effect” (Friedman, 1978) and has been used to argue for some kind of an analogue form of representation, although the only thing that actually follows from the data is the plausibility that all magnitudes may be represented in some common manner (Gallistel, 1990).

There are many other phenomena involving the inspection of visual mental images that appear to parallel those that are found when real scenes are inspected, including the time it takes to judge the similarity of such imagined properties as color (Paivio & te Linde, 1980) or shape (Shepard & Chipman, 1970). Others are discussed in the remainder of this chapter. The parallel between imagery and vision has led a number of people (e.g., Finke, 1980; Shepard, 1978b) to propose that in mental imagery, visual information in memory is fed into the visual system in place of information coming from the eyes. But it should be noted that even if the visual system is involved in mental imagery in this way, it does not in any way speak in favor of the pictorial nature of mental images. As I noted in Chapter 1, the idea that vision involves the construction of an extended image of a scene has been thoroughly discredited and that there is every reason to believe that vision generates symbolic representations. So mental imagery may involve the very same kinds representations as does vision, and yet in neither case need these representations be pictorial, notwithstanding our intuitive impression that there are pictures in our heads. Clearly the picture-theorists wish to make a stronger point than that vision and mental imagery share some mechanisms. They wish to infer from the involvement of the visual system that images are something that can be “seen” which, in turn, would mean that they must be pictorial in nature. The claim that the images function in this way is discussed in Chapter 7, in connection with the use of neuroscience evidence in pursuit of the picture-theory. In what follows I will describe some of the behavioral evidence for the involvement of vision in mental imagery, without specifically raising the

question of whether the evidence addresses the pictorial nature of images.

6.5.1 Interference between imaging and visual perception

One of the earliest sources of objective evidence that persuaded people that imagery involves the visual system is that the task of examining images can be disrupted by a subsidiary visual (or at least spatial) task. (Brooks, 1968) showed that reporting spatial properties from images is susceptible to interference when the response must be given by a spatial method (i.e., pointing) than a by verbal one (i.e., speaking). For example, if subjects are asked to describe the shape of the letter **F** by providing a list of right and left turns one would have to take in traveling around its periphery, their performance is worse if the response is to point to the left or right (or to press left- and right-pointing arrows) than if it is 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 (i.e., both measures d' and β , derived from *Signal Detection Theory*) were affected. Segal and Fusella concluded, “imagery functions as an internal signal which is confused with the external signal” (p 458). This conclusion is, I believe, the correct one to draw. But it does not imply that the same mechanism is involved in the two cases. What it implies, rather, is that interference occurs between two tasks when the same type of representational contents are involved, or the same concepts are deployed. Assume the null hypothesis, that representations in these studies are in a common language of thought, and ask what the representations of visual patterns have in common with representations of mental images? One obvious answer is that they are both about visual patterns. Like sentences about visual patterns, they all involve concepts such as “bright,” “red,” “right angle,” “parallel to” and so on. It is not surprising that two responses requiring the same conceptual vocabulary would interfere. (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.)

6.5.2 Visual illusions induced by superimposing mental images

Other studies suggesting that the visual system may be involved in mental imagery are ones 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 long-lasting orientation-contingent color aftereffect, called the McCollough effect.¹⁰ As I have already suggested, studies that involve projecting images onto visual displays are special in that they provide an opportunity for the visual system to operate on the visual part of the input. In many cases the imagery-induced illusions can be explained simply by noting that in both vision and projected imagery the illusory effect is arguably related to an attention-directing process induced by part of the display. If that is so then imagining that part as projected onto a display may therefore reproduce the same attention manipulation, resulting in the same illusion. Take the following example of the Müller-Lyer effect. When viewed visually, a line appears longer if it has inward

¹⁰ Many of these studies have serious methodological problems that we will not discuss here in detail. For example, a number of investigators have raised questions about possible experimenter demand effects in many of these illusions (Predebon & Wenderoth, 1985; Reisberg & Morris, 1985). Few potential subjects have never seen illusions such as the Müller-Lyer (it is shown in virtually every introductory text in psychology, not to mention children’s books) so even if they do not acknowledge familiarity with the illusion, chances are good that they have some foreknowledge of it. Also 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.

pointing arrowheads at each end (as shown on the right of Figure 2-3). It has been argued that merely imagining the arrowheads produces the same effect. For example, (Bernbaum & Chung, 1981) showed subjects displays such as those illustrated in the top part of Figure 6-9. Subjects were asked to imagine the endpoints of the lines connected 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 found that adding imagined arrowheads also produced the illusion, with the inward-pointing arrows leading to the perception of a shorter line than the outward-pointing arrows. Also (Ohkuma, 1986) found that merely instructing subjects to selectively attend to the inward or outward pointing arrows (as in the bottom of Figure 6-9) produces the same result. I might note that such an effect is not only weak, but it is an ideal candidate for being a classical experimenter-demand effect of the sort discussed by (Predebon & Wenderoth, 1985). But for the sake of argument let us take these results as valid.

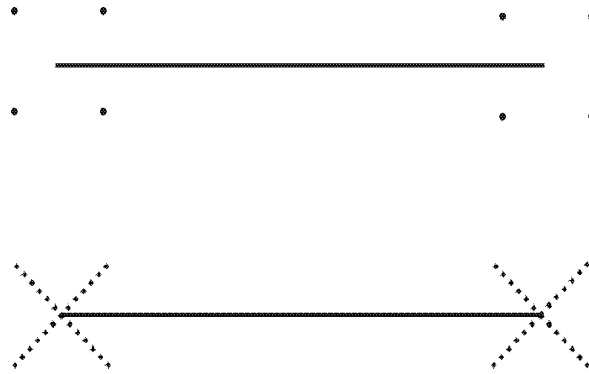


Figure 6-9 Figures used to induce the Müller-Lyer illusion from images. Imagine the end points being connected to the inner or the outer pairs of dots in the top figure (Bernbaum & Chung, 1981) or selectively look at the inward or outward arrows in the bottom figure (reported in Goryo, Robinson, & Wilson, 1984).

Consider first what may be involved in such illusions when the critical parts are actually viewed (as opposed to imagined), using the original Müller-Lyer illusion as our example. Explanations for this 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 direction of gaze. Among the popular explanations that fall into the first category is one due to Richard Gregory (Gregory, 1968), known as the “inappropriate constancy scaling” theory. This theory claims that “Y” type (or inward-pointing) vertices, being generally associated with more distant concave corners of 3D rectilinear structures, are perceived as being further away, and therefore actually larger than they appear. This theory has been subject to a great deal of criticism and is unable to explain a number of findings, including why the illusion is obtained when inducing elements at the ends of the lines are not rectilinear vertices but various sorts of fork-like curves that do not lend themselves to a 3D interpretation (see the review in Nijhawan, 1991). Theories in the second category include ones that attribute the illusion to attention and to mechanisms involved in preparing eye-movement. For example, one theory (Virsu, 1971) claims that the illusion depends on the distance between the vertex and the center of gravity of the arrowhead, and appeals to the tendency to move one's 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 in most cases in which there are context effects on judgments of linear extent. What is important about this from our perspective is that these explanations do not actually appeal to pattern-perception mechanisms and therefore are compatible with attention-based explanations of the illusions. The “envelopes” of the figures in many of these cases can be altered by assigning attention (or visual indexes) to objects or places or regions in the display.

Further evidence that attention can play a central role in these illusions 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, 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. In addition, the relevance of imagery-induction of the Müller-Lyer illusion to the claim that imagery involves the visual system is further cast into doubt when one recognizes that this illusion, like many other imagery-based phenomena, also appears in congenitally blind people (Patterson & Deffenbacher, 1972).

There have been a number of other claims of visual illusions caused (or modified) by mental imagery (e.g., Wallace, 1984a, 1984b). When such image-induced effects on illusions are not due to experimental-demand effects (as they may well be in some cases, where results cannot be replicated under controlled conditions Predebon & Wenderoth, 1985; Reisberg & Morris, 1985) they are all subject to the interpretation that the effect is mediated by the allocation of focal attention. Indeed the attention-mediation of such effects was shown explicitly in the case of ambiguous motion-illusion by (Watanabe & Shimojo, 1998).

6.5.3 Imagined versus perceived motion

Another way to examine the possible involvement of the visual system in imagery is to select some phenomenon known to occur in the early stages of vision and ask whether it occurs in mental imagery. A good candidate is one that involves adaptation to motion, which is known to have a locus in early vision (in fact in visual cortex). 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 (the “waterfall illusion”) and a moving object is seen as moving more slowly (presumably because the motion detection cells in visual cortex have become fatigued). This phenomenon is of special interest to us since the adaptation is known to be retinotopic, and therefore occurs in a retinotopically mapped part of the visual system. Convinced that the visual system is involved in mental imagery, (Gilden, Blake, & Hurst, 1995) set out to show that the motion of an imagined object is similarly 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 (i.e., 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 (i.e., 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 had been habituated or fatigued. But, as Gilden et al. recognized, the effect they observed is exactly what one would expect if, rather than the imagined point moving uniformly 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. We know that people are very good at computing time-to-contact (or arrival time) of a uniformly moving object at a specified location. This is why we are so good at estimating when a baseball will arrive at various critical places (e.g., over the batter's box, at a particular place in the field). What may be going on in imagined motion is that people may simply be using the visual indexing mechanism discussed earlier to pick out one or more marked places (e.g., elements of texture) along the path, and then computing the time-to-contact for each of these places.

We explicitly tested this idea (Pylyshyn & Cohen, 1999) by asking subjects to mentally extrapolate the motion of a small square that disappeared behind an apparently 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 actually appear, as though coming out through a crack in the opaque surface, and then receding back through another crack, and subjects had to indicate whether it had appeared 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 (i.e., 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 (a task at which people are very good; DeLucia & Liddell, 1998) 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.¹¹ This way of describing what is happening does not require the assumption that the visual system is involved, nor does it require the assumption 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.

¹¹ Note that this conclusion applies only to the voluntary movement of a locus of attention over a display, not to apparent motion, which has its locus in early vision. (Shioiri, Cavanagh, Miyamoto, & Yaguchi, 2000) found that observers are good at estimating the location of a moving object along the path of apparent motion. This sort of motion, however, is different in many ways from the motion of an imagined focus of attention, as in mental scanning experiments (e.g., it is not cognitively penetrable – observers cannot change its speed or trajectory at will).

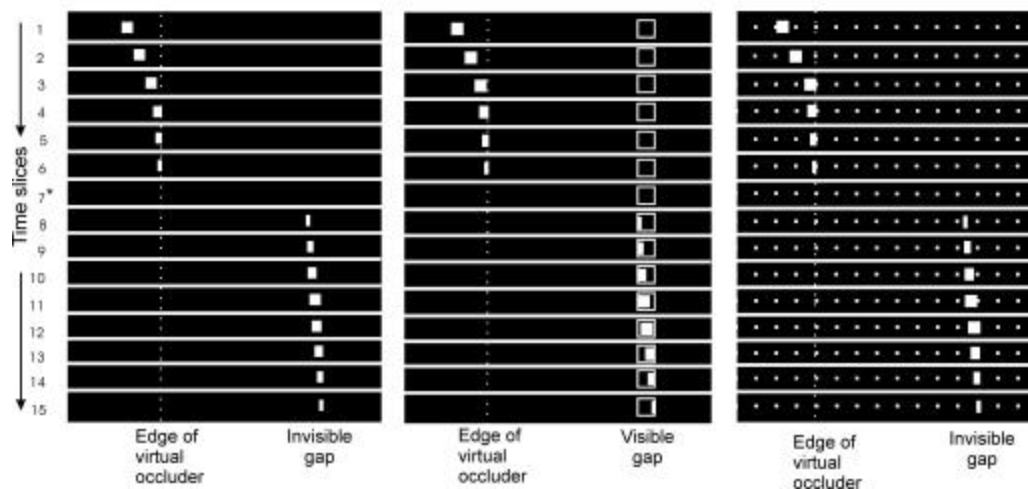


Figure 6-10. Illustration of displays used in (Pylyshyn & Cohen, 1999) to show that smooth motion of imagined objects is poor when no landmarks are visible (panel on left), but better when the gap where objects will reappear was visible (middle panel), or when landmarks were visible along the imagined path (third panel). The results suggest that in imaginal scanning observers are doing a series of time-to-contact computations based on visible locations. (Time slice 7* is longer than shown to allow the object time to get to the location where it will reappear).

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

It is widely held that one of the purposes of mental images is to allow us to discover new visually properties or see new visual interpretations or reconstruals in imaginarily presented information. It would therefore seem important to ask whether there is any evidence for such visual reconstruals. This empirical question turns out to be more difficult to answer univocally than one might have expected, for it is clear that one can draw *some* conclusions by examining images that were not explicitly given in a verbal description. So, for example, if I ask you to imagine a square and then to imagine drawing in both diagonals, it does not seem surprising that you can tell that the diagonals cross or that they form an “X” shape. This does not seem clearly to qualify as an example showing that images are interpreted *visually* since such an obvious conclusion could surely be drawn from a description of a rectangle and of diagonals and also it is a pattern that you have very likely seen before. On the other hand, try the following example without looking ahead for the answer. Imagine two parallelograms, one directly above the other. Connect each vertex of the top figure 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 consistently 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.¹²

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”, a problem to which I have already alluded. Since the everyday (pretheoretical) sense of “vision” clearly involves most of cognition, neither the question of the involvement of vision in imagery nor the question about visual reconstruals can be pursued nontrivially in terms of this broad notion of vision. At the very least we need a

¹² This description accurately, if somewhat unconventionally, describes a Necker Cube.

restricted sense of what is to count as vision or as a visual interpretation. I have argued, in chapters 2 and 3 (as well as in Pylyshyn, 1999) that there is good evidence for an independent visual module, which I referred to as “early vision”. Because early vision is the part of the visual system that is unique to vision and does not involve other more general cognitive processes (such as accessing long-term memory and inference), it would be appropriate to frame experimental questions about reinterpretation of images by examining phenomena that are characteristic of this system. Clearly, deciding whether two intersecting lines form an “X” is not one of these phenomena, nor is judging that when a D is 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 the meaning of such phrases as “looks like an umbrella” (e.g., has a upwardly convex curved top attached below to a central vertical stroke – with or without a curved handle at the bottom). Thus examples such as these simple figures (which were used in Finke, Pinker, & Farah, 1989), cannot decide the question of whether images are depictive or pictorial or, more importantly for present purposes, whether they are *visually* (re)interpreted.

Presenting information verbally and asking people to imagine the pattern being described is one way to get at the question of whether the interpretation can be classed as visual (as in the Necker cube example I cited above). Another way is to present a normally ambiguous pattern and then take it away and ask whether other new visual interpretations occur when the display is no longer there. This case, however, presents some special methodological problems. Not all ambiguities contained in pictures are visual ambiguities and similarly not all reinterpretations are visual reconstructions. For example, the sorts of visual puns embodied in some cartoons (most characteristically in so-called “doodles,” illustrated in Chapter 1, and at URL: <http://www.doodles.com>) do rely on ambiguities, but clearly not on ones that concern different visual organizations analyzed by the early visual system. 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 question we would have to control for certain alternative explanations of apparently visual 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 alternative view that many writers have preferred (Casey, 1976; Fodor, 1981). Another possibility might be that the observer actually computes both alternatives, but only reports one. This sort of simultaneous computing of two readings, which only last for a brief time, has been frequently reported in the case of sentence ambiguities. With lexically ambiguous sentences, it was shown that both interpretations are briefly available even though the person was aware of (and able to recall) only one of the senses. For example, (Swinney, 1979) showed that *both* senses of an ambiguous word such as “bug” (as in “There were many flies spiders and other *bugs* in the room” as opposed to “There were microphones, tape recorders and other *bugs* in the room”) primed associated words for a short time after they were heard in these two contextually disambiguated sentences (i.e. both “insect” and “spy” were read more quickly within a second or so of hearing “bug”). Similarly, in chapter 2 (as well as in Pylyshyn, 1999), I argued that notwithstanding the encapsulated nature of early vision, one of the ways that the cognitive system might still be able to affect the visual interpretation of ambiguous figures is if both possible interpretations were initially available so that the cognitive system could select among them based on their plausibility. Despite the methodological difficulties in obtaining a clear answer to the question of whether mental images allow visual reconstructions, a number of studies have provided converging evidence suggesting that no visual reconstructions occur from mental images.

(Chambers & Reisberg, 1985) was 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) have taken issue with these findings, citing their own experiments involving operations over images (e.g., the D-J umbrella examples that were mentioned briefly in section 6.1), but as I suggested above it is dubious that the reinterpretation of the superposition of such simple familiar figures should be counted as a *visual* reinterpretation. Moreover, even if the interpretations studied by Finke et al. were considered visual interpretations, there remains the serious problem of explaining why clear cases of visual interpretations, such as those studied by Chambers and Reisberg, do not occur with images.

Mary Peterson undertook a series of detailed explorations of the question of whether mental images can be ambiguous. Peterson and her colleagues (Peterson, 1993; Peterson, Kihlstrom, Rose, & Glisky, 1992) argued that certain kinds of reconstruals of mental images do take place. They first distinguished different types of image reinterpretations. In particular they distinguished what they called *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 they called *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). I 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 systematically confirmed by Slezak, 1995, using quite different displays). Peterson and her colleagues 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 independent of the question of which mechanisms are used in processing it. What I am concerned about is whether the format of images is such that their interpretation and/or reinterpretation involves the specifically visual (i.e. the early vision) system as opposed to the general inference system. The crucial question, therefore, is how Peterson's findings on reinterpreting mental images compare with the reinterpretations observed with ambiguous visual stimuli. The answer appears to be that even when reinterpretations 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 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 (as illustrated in Figure 6-11), whereas 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 discussed in the next paragraph). 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 early vision is sensitive to such cognitive influences – as we saw in section 2.5.2 – although later stages clearly are.

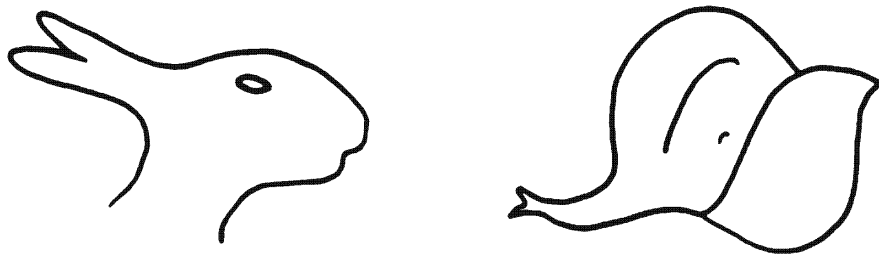


Figure 6-11. Ambiguous figures used by (Peterson, 1993). When these figures are perceived visually, the percept tends to be bistable and changes categories in a radical way (e.g., duck-rabbit on the left and elephant-snail on the right). When these are reconstructed from a mental image they tend to be reinterpreted in a wide variety of other ways, suggesting that local cues are being used to provide plausible reinterpretations, rather than re-perceptions. 9These figures are stylized versions of well-known illusions introduced by (Jastrow, 1900) and by (Fisher, 1976), respectively.)

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 reconstructions is not a perceptual (re)interpretation of an internally generated picture, but something else, perhaps the sort of inference and memory-lookup based on shape properties 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 the other figures used, 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 be made from among them based on a search for a meaningful familiar shape 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 “reconstructions” 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.

The clearest source of evidence I am aware of that bears on the question of whether mental images can be reconstructed is provided by Peter Slezak (Slezak, 1991, 1992, 1995). Slezak asked subjects to memorize pictures such as those shown in Figure 6-12. He then 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. Subjects had all the relevant information since they could draw the rotated figures from memory, and when they did, they usually recognized the rotated shape from their drawing! What is special about these examples is that the resulting 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 are not familiar, nor do they contain unique “landmark” features (which can be used to recognize the rotated pattern – see Hochberg & Gellman, 1977), and their appearance after the rotation could not easily be inferred from their description.



Figure 6-12. 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 the figures). What do you see? Now try it by actually rotating the page. (from Slezak, 1995)

6.5.5 What about the *experience* of visualizing?

It may well be that the most persuasive reason for believing that mental imagery involves an inner perception is the subjective one: Mental imagery is accompanied by an experience very similar to that of seeing. As I remarked at the beginning of this chapter, this sort of phenomenal experience is very difficult to ignore. Yet studies of visual perception demonstrate that introspection is highly suspect source of evidence about the nature of the underlying representation, because it trades on the failure to distinguish between the content of our experience (what we are imagining) and the causal entities in our mind/brain that are responsible for the way things appear in our imagination. There is the very strong temptation to reify the image content – to make what philosophers call the “intentional fallacy” (and that Titchener called the “stimulus error”) of attributing to a mental state the properties of a world that it represents.

But the experience of imaging is not an experience that reveals the form of the image; rather it reveals the content the image. The experience of the image tells us what it is an image *of*. Because of this it is plausible that both vision and imagery lead to the same kind of experience because both are mediated by similar internal states. In fact, a common view, made explicit by Roger Shepard (Shepard, 1978b), is that imagining X is nothing but the activation of mental states normally associated with the visual perception of X, though without X being present. Shepard notes that because the image is not external, the experimenter cannot discover its shape or other visual properties. But if we *could* externalize it, by mapping from a brain state back to a stimulus, which could have been its distal cause, it would look like a scene – as illustrated by Shepard in Figure 6-13 and Figure 6-14.

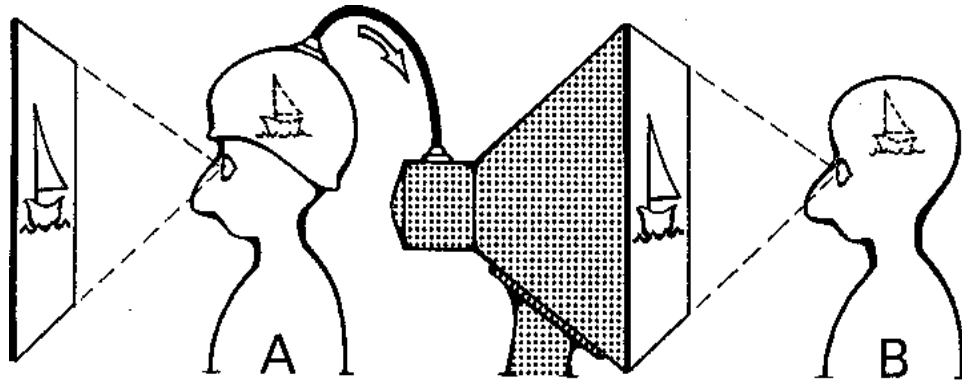


Figure 6-13. Illustration of a hypothetical mechanism (in a “thought experiment”) that externalizes the brain states involved in visual perception (by A) so another person (B) might have the same visual experience without the presence of the original visual stimulus (from Shepard, 1978b).

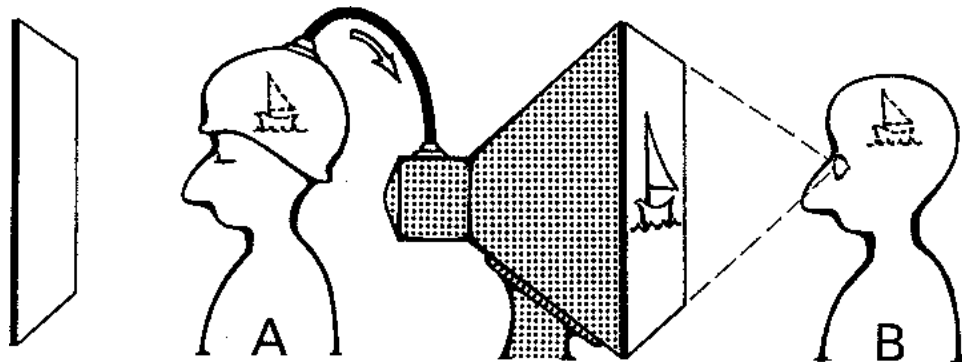


Figure 6-14. Illustration of how the apparatus introduced in the “thought experiment” of Figure 6-13 above, could also be used to externalization a mental image as a visual display (from Shepard, 1978b).

Shepard recognized that even if the experience of mental imagery arises from the same brain states as occur in visual perception, this would not entail that the internal state is underwritten by a pictorial, as opposed to a symbolic structures in the brain. However, the interpretation that Shepard presents in the above two figures does assume that a vision state and an imagery state carry sufficiently similar information that one could, at least in principal, put another person’s brain into the same state as that of a person having a mental image, by providing the second person with appropriate *visual* stimulation. The proposal of a possible (though fanciful) way to externalize the visual experience of a person imagining some situation, assumes that a mental image and an actual display can be *informationally equivalent*. This turns out to be a very strong assumption, and one that is very unlikely to be true in general (as Shepard himself recognized – see note 66).

The situation shown in the above figures is clearly not possible in principal. In particular, B’s brain could not, even in principal, be put into a state corresponding to A’s having a mental image, by providing inputs *through B’s eyes*, using a display. Consider, for example, that in Figure 6-13, A is able to examine the figure freely with eye movements as well as with covert movements of attention while B need not be scanning his display in synchrony or even scanning it at all. There is good reason to believe that without making the same

voluntary eye movements and attentional scannings, A and B would not have the same experience (see, for example, O'Regan & Noë, 2002). B is in a position to examine the display freely, to scan it with eye movements and to selectively attend to parts of it. This assumes that the visual information about the scene in A's brain (which by assumption has been faithfully mapped onto the display that B sees) contains *all* the panoramic information needed to allow B to examine, scan and interpret it – an assumption that we already found to be untenable since the information about a scene that is actually encoded by A is a very small part of the information that could be encoded given sufficient time. But B would have to be shown *all* the potential information in order to be able to scan the scene and thus to have the same experience as A (see section 1.4, as well as section 6.5.4 above).

The problem is that the experience of “seeing” is not informationally equivalent to *any* display because it must include potential as well as actual inputs. Part of the reason for this is that A is able to use visual indexes or some similar deictic mechanism to keep in contact with the display in order to access additional information as needed, thus ensuring that more of A's display is potentially available than is encoded or noticed at any point in time, whereas this potential does not apply to the display shown to B since that only includes the current state information. As I suggested in chapter 5, it is this capability to return to the display at will that confers the apparent panoramic and detailed pictorial feel to our perception of a scene.

Even more important is the fact that what A experiences is an *interpretation* of the stimulus; it is, for example, one view of an ambiguous display. Since what is presented on B's display is pictorial, and therefore uninterpreted, B is free to construe it differently from the way A construed it, even if it somehow accurately maps the contents of A's visual cortex. For example, suppose in Figure 6-13 A is looking at the ambiguous Necker Cube (e.g., such as one of the Necker Cubes shown in Figure 1-7). A will see it in one of its two construals, so this will be the contents of his phenomenological experience and, by assumption, of the state of his brain. But what will appear on B's display in that case? One cannot construct a pictorial image of the cube that already has one of its construals preselected (even if one were allowed to label its parts). So then B is free to see a different construal of the Necker Cube than A does, contradicting the assumption that the display has the possibility of conveying the phenomenological experience from A's brain to B's display. The situation in Figure 6-14 has the same problem: When A imagines a Necker Cube he imagines it with one or the other construal (recall that Peterson found no cases in which people could see reversals of the Necker Cube in their image). Thus, again, B is in a position to have a different perceptual experience when the visual information is projected from A's brain to B's screen.

The problem arises because Shepard was attempting to “externalize” the contents of the experience of “seeing” (either seeing a visual scene or a seeing mental image), but the experience corresponds to having in mind an interpretation, not of having a raw unanalyzed array of light and dark colored regions.¹³ As J.J. Gibson and many others before him have stressed, we do not experience patches of light and color, one experiences seeing people and tables and chairs – the ordinary familiar things that populate one's world, the recognition of which requires that we consult our long-term memory. That may be why the experience of seeing and of having a mental image very likely corresponds to activity at a high level of the visual process, after the stimulus is recognized and interpreted as something familiar. In fact, a number of

¹³ Roger Shepard recognizes the inadequacy of the “externalization” view entailed by his Figure 6-13 and Figure 6-14, but he then goes on to say that this “deep structure” aspect of the percept is being overplayed because most figures are not ambiguous and most people see the same things when presented with the same displays. But this concession misses the point since even if most people saw the same thing, it does not entail that the experience of having a mental image can be rendered on a display. It is in the nature of real displays that they are uninterpreted and therefore boundlessly ambiguous. Even if you could somehow create images that were unambiguous for most people most of the time, they would still need to be interpreted and the interpretation would not be equivalent to a display. This can be seen clearly in the case of experienced mental images because mental images are conceptual: one has a mental image under one's intended interpretation. One does not need to provide an interpretation of one's own image and indeed one generally can't provide a new *visual* interpretation of one's mental image. This and related points will be discussed in chapter 7.

neuroscientists have suggested that the locus of awareness (assuming that awareness even has a brain *locus*) occurs much higher in the visual system than primary visual cortex (Crick & Koch, 1995; Stoerig, 1996). So if the experience of seeing or of imagining is to be transmitted from A to B, then at the very least, it would require that the state of a much larger part of A's cortex would have to be mapped onto B's cortex.

In addition to these problems about conscious contents, it is also unlikely that the distinction between information processing episodes of which we are conscious and those of which we have no awareness marks a scientific natural kind; the kind of demarcation that can help to focus our attention on the relevant causal variables. It could turn out that imagery is, after all, a special form of information processing that uses a special modular system, but that it only overlaps partially with episodes of which we are consciously aware as a kind of seeing. It *could* turn out that the correct analysis of the type of information processing involved in mental imagery does not distinguish between episodes of which we have a sensory-like awareness and those of which we do not. In other words, it could turn out that there is a theoretically interesting and distinct form of information processing which is sometimes accompanied by the phenomenology we associate with imagery and sometimes not. Already there has been talk of "unconscious imagery" and there is evidence that most experimental results obtained in studying imagery (such as the mental scanning results, the image size results, the mental rotation results, and others discussed earlier in this chapter) can be obtained with congenitally blind subjects (see, for example, Barolo, Masini, & Antonietti, 1990; Cornoldi, Bertuccelli, Rocchi, & Sbrana, 1993; Cornoldi, Calore, & Pra-Baldi, 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) as well as with subjects who profess little or no experience of mental imagery (see Chapter 7 for more on the dissociation between imagery and vision). It is also significant that the status of reports of vividness of images is problematic. Subjective vividness does not appear to correlate with performance in a great many imagery tasks. For example, it has been reported that the ability to recall visual experiences is either uncorrelated (Berger & Gaunitz, 1977) or even negatively correlated with vividness of the experience (Chara & Hamm, 1989), although the confidence level of the reports is correlated with vividness (McKelvie, 1994) as is performance on tasks such as inducing visual illusions by superimposing imagined patterns over perceived ones (see section 6.5.2). Other tasks show differences between vivid and nonvivid imagers that are hard to interpret in terms of their use of mental images (see, e.g., Wallace, 1991).

In assessing the evidence from phenomenology one needs to seek objective corroborating evidence for informational differences between mental episodes accompanied by the experience of seeing and mental episodes that are accompanied by some other experience (or no conscious experience at all). A great many studies suggest that unconscious states play the same role in cognition, as do conscious states (e.g., stimuli of which we have no conscious awareness appear to influence perception and attention the same way as stimuli of which we are aware, Merikle, Smilek, & Eastwood, 2001). But there is also some different information processing and neurophysiological correlates of episodes accompanied by awareness (Dehaene & Naccache, 2001; Driver & Vuilleumier, 2001; Kanwisher, 2001). What I have argued, here and elsewhere, is just that we are not entitled to take the content of our experience as reflecting, in any direct way, the nature of the information processing activity (what Pessoa, Thompson, & Noë, 1998, call the "analytical isomorphism" assumption). In particular, the evidence does not entitle us to conclude that episodes that we experience as *seeing in one's mind's eye* involve examining uninterpreted, spatially-displayed depictive representations (i.e., pictures) using the early visual system.

Finally, cognitive science has no idea what to make of the subjective impression that we are experiencing a “seeing” or a “hearing” episode, since this is tied up with the deeply mysterious questions of consciousness. It is always very tempting to view our experience *as such* as playing a causal role – as having causal powers. We say such things as that we scan our image or zoom in to see a small detail in it or that we remember a certain emotional crisis by thinking of it in a certain way (e.g., as having a spiral shape). I am sure that there is a sense in which all these are true reports. But this way of putting it suggests that how a cognitive episode is experienced *causes* a certain observable consequence, rather than the cause being some underlying physical state. Taken literally, such claims entail a dualist (or at least an interactionist) metaphysics. They claim that there can be causes that are not physical (i.e., that do not follow natural physical law). While that is a coherent position one could take, it is not one that most writers on mental imagery would want to defend. What we would prefer to defend is the notion that underlying a certain type of experience are certain physical/biological properties and it is these properties that are the causes of the behavior. A particular experience, say the experience of “seeing in the mind’s eye”, is not a possible natural property, so the statements that contain references to such experiences as causes are not literally what their authors want to claim. But if it’s not literally the case that I behave in certain ways because of how I experience some mental episode, then I need a different way to talk about the causal sequence, a way that will not attribute causal properties to experiences the way that everyday discourse sometimes does; a way that does not mention that I do something or think something because it feels like I am seeing a picture in my mind’s eye.

7. SEEING WITH THE *MIND'S EYE* 2: SEARCHING FOR A SPATIAL DISPLAY IN THE BRAIN

One of the most widely accepted claims about mental images is that they are “spatial”; and among the most frequently cited demonstrations of this claim are the mental scanning experiments discussed in section 6.3.2.1. There I argued that it is unwarranted to conclude that because it takes longer to scan attention across greater imagined distances that images must “preserve” metric spatial information (Kosslyn et al., 1978) – at least not if by “preserving” distances one means that images *have* spatial properties such as size and distance, rather than simply encoding such properties (in *any* way). Yet the alleged spatiality of mental images is central to the claim that images are *depictive* and thus unlike any possible “language of thought.”

The claim that images have spatial properties comports with one’s intuitions about mental images, but it raises some interesting and subtle questions about what it means to *have* spatial properties, and whether there could be any sense to this notion apart from the literal claim that images are written on some physical surface, presumable on some surface in the brain. In what follows I will examine the claim that images are spatial and then examine the recent evidence that has been presented and interpreted as bearing on this question.

7.1 Real and “functional” space

One of the more seductive claims about images made by Kosslyn (e.g., in the quotation cited in Chapter 6, section 6.4) and echoed many other writers (e.g., Denis & Kosslyn, 1999; Prinz, 2002; Tye, 1991), is that images are laid out spatially in a *functional*, rather than in a physical space – where the notion of a “functional” space is introduced by example, as something like a matrix or array data structure in a computer. This is a seductive idea because it appears to allow us to claim that images are spatial without also committing us to claiming that they are actually laid out in real space in the brain or in some other physical medium. It also appears to give some meaning to the claim that images somehow incorporate (or “preserve”) distances as well as sizes. Because the idea that images are somehow spatial has been so influential in theorizing about mental imagery, I will devote this chapter to a discussion of this claim and to some of the recent evidence cited in support of it. Before going into the evidence, however, I consider what it might mean to claim that images are spatial. Then I will argue that the appeal to a functional space is actually empty and merely restates the phenomena we are trying to explain.

The problem with the “functional space” proposal is that functional spaces do not intrinsically have any particular properties. Being “functional” they are not subject to any natural laws and therefore can be assumed to have whatever properties are needed in order to account for the experimental data. Because a functional space has no intrinsic properties, any properties it has are ones that are stipulated or extrinsically assumed, so it can accommodate any findings whatever. Or, to put it more positively, if the extrinsic assumptions provide any explanatory advantage at all, they can equally well be adjoined to a theory that assumes *any* form of representation, not just a pictorial one. They can, for example, be added to a theory that claims that the form of information underlying images is that of a “language of thought.” Despite this clear inadequacy of the functional space proposal, the idea remains widely accepted in psychology, and even in some quarters of philosophy. The reason for this seductive quality of the proposal is itself quite revealing and so I will devote part of this chapter to discussing how the functional space idea connects with the literal space

(the display-in-the-head) hypothesis, and how it maintains a very strong pull on contemporary research on mental imagery which itself has led to certain lines of investigation of mental imagery, particularly in neuroscience.

In the computational model of mental imagery described in (Kosslyn et al., 1979), the inner “screen” on which images are “displayed” is a matrix of elements – a computer data structure that corresponds to a two-dimensional matrix. It functions like a two-dimensional display in several respects. Graphical elements, such as points, lines, contours, regions, and other components of figures are written on the matrix by filling designated cells which correspond to places on the image, defined in terms of quantized Cartesian coordinates (pairs of matrix indexes). The principles for generating figures are explicitly stated, as are principles for reading off figural properties (at least in principle, if not in an actual working system). For example, in order to examine a certain part of a figure after examining another part, the system has to move a locus of processing (corresponding to its focal attention) through intermediate points along the path from one figure part to the other. This is done by incrementing the (x and/or y) indexes by one. The matrix itself appears *intrinsically* to contain unfilled places that serve as potential locations of figural elements, so that empty places appear to be explicitly represented. As Kosslyn claimed in the earlier quotation, in this kind of 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. ... not only is the shape of the represented parts immediately available to appropriate processes, but so is the shape of the empty space...” In other words, empty places are explicitly displayed, along with filled places, so that regions between contours are formed and displayed as a natural consequence of displaying contours. This form of representation appears to embody the criteria for being a depictive representation as described in the quotation. In this way the model appears to (at least in principle) account in a natural way for much of the relevant data, as well as the qualitative aspects of our experience of imagining, so what else could one ask for?

As we saw when we discussed the mental scanning results (and the mental color-mixing example) in the last chapter (section 6.3), in order to provide an explanation it is not enough that a model exhibit certain patterns of observed behaviors. We still need to say *why* these behaviors arise. As with all the examples I discussed earlier (including the mystery code-box example), it makes a great deal of difference whether a particular generalization holds because (a) the system uses its representations of the world, along with its reasoning capacity, to simulate some behavior it believes would occur in the imagined situation, or because (b) the principle that determines the behavior of the system is inherent in the mechanism or medium or architecture of the system that displays and examines images. So we need to ask which of these is being claimed in the case of the matrix (or any other “functional space”) model of the image? The answer is crucial to whether the model provides an explanation of the experimental findings or merely summarizes them, the way a table or other descriptive data summary might do.

Here the image theorist is caught on the horns of a dilemma. On the one hand the system (say the one that uses a matrix data structure) might be viewed as a model of the architecture (or, equivalently, the format or the medium) involved in mental imagery. If that is the case, its properties ought to remain fixed and they ought to explain the principles by which the imagery system works (and perhaps even provide some suggestions for how it might be implemented in the brain). This proposal (which we might call a “model of the architecture of mental imagery”) can be subjected to empirical scrutiny, using, among other techniques, the criterion of cognitive penetrability. The architectural version of the functional space proposal entails that

the observed properties will not change with changes in the person's beliefs about the task. On the other hand, the matrix model might be thought of as simply summarizing the pattern of people's behavior in particular circumstances (e.g., the increase in time to switch attention to more distant places in the image, or the increase in time to report details from a "smaller" image), with no commitment as to *why* this pattern of behavior arises. In this purely descriptive view of the imagery model, the matrix is compatible with the pattern being due to what subjects understood the experimental task to be (e.g., that the task is to simulate as closely as possible what would happen in a situation in which they were looking at a real display) and to their tacit knowledge of what would happen if they were to see certain things occur. Summarizing such a pattern of behavior is a perfectly legitimate goal for a theory, although it only meets the criterion that Chomsky calls "descriptive adequacy" as opposed to an "explanatory adequacy". Even in this case *some* general type of mechanism might have to be assumed, for example a mechanism capable of representing beliefs, drawing inferences from them, generating representations of some intermediate states of the simulation, as well as estimates of the relative times involved. What would not be relevant, however, are assumptions about the format in which the images are stored, displayed and accessed, since those properties would be doing no work here (the test being that the operative assumptions could be added to any model, regardless of the format it used for representing the relevant beliefs). It is clear from the writings of picture-theorists that this is *not* what they have in mind; rather they take their CRT model (or what I called their "picture theory") to be what we have called a *model of the architecture* of mental imagery (certainly this is explicitly claimed in Kosslyn, 1981).

Consider first the "matrix model" (Kosslyn, 1994) viewed as an architectural model. For now let us see why such a model appears intuitively attractive and what in fact it assumes about the mental architecture underlying the use of mental imagery. Later I will consider whether the empirical facts are consistent with such a model. The conclusion I will reach is that a matrix is attractive precisely because it *can* be (and typically is) viewed as a computer simulation of a real 2D surface, which is precisely the *literal* picture-in-the-head assumption one was trying to avoid by talking of "functional space." Consider some reasons why a matrix appears to be spatial. It appears to have two dimensions, to explicitly *have* distances (if we identify distance with the number of cells lying between any two particular cells), as opposed to merely representing them (say in some numerical code), and it appears to explicitly *have* empty spaces. It also seems that it can be "scanned" from one place to another, since there is a natural sense of the topological property of being "adjacent," so that searching for an object by passing through intervening empty places makes sense. The trouble is that all these spatial notions – adjacency, distance, unfilled places and scanning – are properties of a way of thinking about a matrix; they are not properties of the matrix as a data structure. Looked at as a data structure, there is nothing "adjacent" between cell (25, 43) and cell (25, 44) except when we view them as points in a real space, as opposed to two quite discrete registers in the computer, which is what they are. What we think of quite naturally as two *locations* in space are nothing but two discrete symbols or register names (which in the case of a computer, we refer to as "addresses," or even "locations" in memory, thus inviting a spatial metaphor). What makes them appear to be locations is not inherent in their being indexes of cells in a matrix data structure; it is a property that is tacitly introduced by our way of thinking about the data structure. For example, according to Kosslyn, "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)." But that isn't the way a matrix has to be *used unless it is being used to simulate a real spatial surface or display* and then it is not the matrix but the thing being simulated that has the spatial properties. In a matrix, one does not get to a particular cell by "moving" there

from an “adjacent” cell – one simply retrieves the contents of the cell by its name (names happen to be pairs of numerals, but that is of no consequence – the pairs are converted to a single symbolic address anyway). Of course when it is used as a model of space, accessing registers is constrained in the way it is constrained in physical space, viz, one gets to a cell by accessing a sequence of cells defined as being between the currently accessed cell and the desired cell. But in that case it is this extrinsically stipulated constraint that is responsible for this way of accessing information, not the fact that it is a matrix; and of course the reason the extrinsic constraint is stipulated is because the theorist implicitly believes that the mental image is actually written on a physical surface, not a “functional” surface! This is the only possible motivation for what would otherwise be a totally arbitrary constraint.

Viewed in this way, the notion of a functional space has no explanatory role because a functional space has whatever properties we require it to have. If we wish the functional space to have the properties of real space why then we arrange for the relevant properties to be manifested. But then we cannot appeal to something we call a functional space to explain why it has those properties: It has those properties because we stipulated them. The spatial character of a matrix (or any other implementation of “functional space”) is something that is stipulated or assumed *in addition* to the computational properties of the data structure. The critical fact here is that such an extrinsic postulate could be applied to *any* theory of mental images, including one that assumes that images are sentence-like symbolic structures. Indeed, there is nothing to prevent us from modeling an image as a set of sentences in some logical calculus and then, in addition, stipulating that to go from one place to another (however places are referenced, including the use of numerals as names) requires that we pass through “adjacent” places, where adjacency is defined in terms of the ordinal properties (or even metrical properties, if the evidence merits that assumption) of place names. If this sounds *ad hoc*, that’s because it *is*! But the assumptions are no more *ad hoc* than they are when applied to a matrix formalism. For example, in a matrix data structure it is equally natural to go from one “place” to *any* other place (think of places as having names, you can go from a place named Q1 to a place named Q2 just by issuing those two names to the function that accesses cells in a matrix). If you want to restrict the movement you have to do it by assuming a constraint that is extrinsic to the form of the matrix. It is only if you view the matrix as a representation of a real (physical) two-dimensional display that you get a notion of “adjacent cell,” and it is only if you confine operations on names to ones that move to adjacent cells defined in this manner that you get a notion of scanning. The same is true of the claim that matrices represent empty places. An empty place is just a variable name with no value: Any form of computational model can make the assumption that there are names for unoccupied places (in fact you don’t even have to assume that these places exist prior to your making an inquiry about their contents – as they don’t in some computer implementations of sparse matrices).

The point here is not that a matrix representation is wrong. It’s just that it is neutral with respect to the question of whether it is supposed to model intrinsic properties of mental images – the property that is responsible for the many experimental results such as scanning – or whether it is supposed to be a way of summarizing the data which actually arise from people’s knowledge of how things happen in the world (i.e., that when using mental images, people usually act as though they are observing things in the world). When appeal to the matrix appears to provide a natural and principled account of the spatial imagery phenomena (such as scanning and the effect of image size) it is invariably because the theorist is tacitly viewing it as a representation of a two-dimensional physical surface. In other words, the theory is principled if and only if it assumes that mental images are written on a physical surface. A real surface *has* the relevant properties because physical laws apply to it. For example, on a real physical surface it takes time $t = d/v$ to traverse a

distance d at velocity v and this is guaranteed by a law of nature. In a “functional” space the relation between d , v , and t is indeterminate: It has to be stipulated in terms of extrinsic constraints. In a real physical surface the relative 2D distances among points remains fixed over certain kinds of transformations (e.g., rotation, translation, change of size) and the Euclidean axioms always hold of distances between them. In particular, distances obey metrical axioms; for example the triangle inequality must hold of the distances d among three elements x , y , and z so that $d(x,y) + d(y,z) = d(x,z)$. In a functional space none of these constraints is required to hold by virtue of the format or intrinsic nature of the representation (unless, once again, functional space is just a gloss for real space).

Of course it is possible, in principle, to implement a model of space in another (analogue) medium by mapping sets of properties of space onto certain properties of the model medium (that’s what an analogue computer typically does). For example, one might use the relationship among voltage, resistance and current flow (specified by Ohm’s Law) to model the relationship between distance, speed and time, respectively. Although in this case the properties of the model are different from the properties being modeled, they are nonetheless intrinsic properties – certain properties and relations in the model hold because of the physics of the analogue medium. A system that follows Ohm’s Law constitutes an analogue model of a system consisting of the three properties involved in mental scanning (distance, speed and time). That part is easy because it involves the relationship among only three variables. But it would be surprising if *all* properties of space could be modeled by something other than space itself. Notice that the above analogue model of distance, speed and time would fail to model lots of other aspects of space and motion such as the principles that hold among distances and relative directions, or distances and areas, since Ohm’s Law provides no natural quantity corresponding to direction or area or even the time course of change of voltage with respect to current which ought to map onto the distance traveled over time. One could, of course, define such quantities as area or trajectory in terms of some functions over time, distance and current flow – but these would have to be calculated, as opposed to being measured directly in the model. It would also fail to model such interesting relationships as those embodied in Pythagoras’ Theorem, the invariance of some properties or relations (like area or relative distances or relative angles) over certain transformations (such as rotation). It would clearly not be able to model such assumed image properties as the visibility of an object’s features as a function of the size of the object. There are an unlimited number of such properties and relationships, which is why it’s hard to think of any system of variables other than those intrinsic to space and time that could model them all. The problem of finding a non-Cartesian model of Euclidean space is precisely what concerned Jean Nicod. His beginnings on this problem (Nicod, 1970), though brilliant in itself, did not go very far towards providing possible models of space that could be of use in understanding mental imagery.

But even though such considerations are relevant to the attempt to make sense of the intuitive picture-theories of mental images, this entire discussion of analogue representation is beside the point when it comes to explaining most mental imagery phenomena. The real question is not how to implement an analogue representation of space, but whether one should model the apparent spatial character of images in terms of the architecture of mental imagery. My purpose in belaboring the issue of what constraints are imposed by the format and what constraints are imposed by extrinsic stipulations, is simply to clarify the role that something like a matrix or other “functional space” *could* play, and thereby to set the stage for the real issue, which is the empirical one. I have already described some of the empirical findings concerning mental scanning (e.g., our own studies of mental scanning when the task was understood as imagining lights that go off in one place and simultaneously go on at another place, or in which the task focused on aspects of retrieving information from an image that play down the question of how one gets from one place to another).

Such findings show that no analogue representation could explain the results since the results depend on observers' beliefs. The same goes for the finding that it takes longer to answer questions about small visual details when the object is viewed as "large" than when it is viewed as "small." The way the data turn out in these cases shows that the mind does not work as though the imagery architecture imposes constraints like those you would expect of a real spatial display, at least not in the case of these particular phenomena.

But how about other properties of spatial representations; are the properties of represented space cognitively penetrable? Can we imagine space not having the properties it has? Kosslyn has claimed that the spatial property of the image display constrains what we can imagine. He says (Kosslyn et al., 1979), p549], "We predict that this component will not allow cognitive penetration: that a person's knowledge, beliefs, intentions, and so on, will not alter the spatial structure that we believe the display has. Thus we predict that a person cannot at will make his surface display four dimensional or non-Euclidean..." But as I remarked at the beginning of Chapter 6, the more likely reason why people cannot image a four-dimensional or non-Euclidean world is that they do not know what such a world would *look like* and knowing what it would look like is essential to being able to carry out the task we call "imaging." According to this view, not knowing where the edges and shadows will fall is alone sufficient reason for failing to image it without any need to invoke properties of a "surface 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 as quickly or as slowly as we wish, with or without visiting intermediate (empty or filled) places. In fact, as we saw earlier (section 6.5.3), some of our recent work provides reason to doubt that we *can* smoothly scan a mental image (or even a real stationary scene) or visit more than a few places along the way, even if we wanted to. Rather it looks as though the best we can do visually is estimate how long it would take an object traveling at a given speed to move between a few places we can literally see, and the best we can do in our mental image is to imagine *that an object is moving*, but not to imagine the object as being at a succession of intermediate places along the way (at least not more for than a few such places). Since the spatial nature of mental images is such a well-entrenched belief, I will devote the next section to a discussion of what it would mean for images to be spatial, and also to a brief discussion two potential reasons why in certain situations images may actually exhibit some nontrivial spatial properties without themselves being laid out on a spatial surface.

7.2 Why do we think that images are *spatial*?

7.2.1 Physical properties of mental states: crossing levels of explanation

In chapter 1 I mentioned the strong temptation to reify experiential properties like color or shape or duration by attributing these properties to mental representations. This tempting mistake is the one that slips from saying that you have an image of a scene and that the scene has property P, to the conclusion that the mental event or the formal object in your head itself *has* property P. The temptation is to equivocate between the two ambiguous senses of the phrase "an image of X with property P"; namely "an image of (X which has property P)" and "(an image of X) which has property P." In the first case it is X, a thing in the world, that has the property, whereas in the second case it is the image of X, a thing in the head, that is claimed to have the property. In referring to one's representational states (e.g., one's thoughts or images) the relevant properties are always properties of the thing one is imagining, not of the object that stands in for it – the representation itself. The degree of the temptation to reify properties varies with the particular property. It is

pretty easy to see that an image of a cat does not have to be furry, an image of an orange need not be colored orange, or that an image of one object on top of another need not be made up of 2 parts, one of which is on top of the other. Nonetheless, it is a bit harder to accept that an image of a big thing need not be large (or larger than an image of a small thing), and harder still to appreciate that imagining a long-lasting event need not last long.

I once wrote a brief commentary (Pylyshyn, 1979a) in which I reflected on how freely we speak of the beginning, end and duration of mental events – for example, of the representations produced by perception or by imagination. I wondered why we were not disposed to speak of other physical properties of mental events, such as their temperature or mass or color. After all, these are all legitimate physical properties: If the result of perception is a physical state or event, which few people would deny, then why do we find it odd to ask what color or weight or temperature a percept is, somewhat less odd to ask how big it is or where it is, but are quite comfortable about asking how long it took. The commentary did not elicit any response – people found it puzzling why I should even raise such a question when it is obvious that mental events have durations but not temperatures. But this should have been a puzzle. The reason we do not speak of the physical properties of mental events is not because we are closet dualists who believe that mental things are in a different realm than physical things, but because mental event *types* need not correspond to physical event *types*. To speak of an event type is to refer to an entire class of events among which we make no distinction for certain purposes. All the instances of the letter “a” on this page are considered equivalent for the purposes of reading, even though they may differ in size, font, and in other ways of interest to a typesetter – and they certainly differ in their location. Thus we speak of the letter *type* “a” of which there are many letter *tokens* or instances of “a” on the page that are distinguishable in many different ways. There is no answer to the question; What size is the letter “a,” because it can be any size and still be the letter “a.” But a mental event type has many event tokens each of which could have arbitrarily different physical properties – at least they are not required to have any fixed property merely by virtue of being a member of a particular event type. When we ask whether a mental event (such as imagining a particular melody being played) has a particular duration, we are asking about the event-type, not the event-token. There need be no answer to the question: “What is the temperature of an imagining of the singing of *Happy Birthday*,” though there is an answer to the question, “How many words are there in an imagined singing of *Happy Birthday*” (at least according to a certain interpretation of what it means to “imagine X”, namely that to imagine X one must imagine, as best one can, each part of X in the right sequence and with roughly the right relative duration).

To make the idea of an event-type more concrete, consider the parallel case in computing. The event of computing a certain function (say a square root function) is realized physically in a computer; therefore it is a physical event – no doubt about that. Yet *where* it is carried out it is not relevant to its being that particular function, in part because it may occur anywhere in the machine and in fact will occur over widely dispersed locations from occasion to occasion. There is no physical location corresponding to the function *as such*, even though each token occurrence of such a function does take place somewhere or other. This is what we mean by an event type: it is what all occasions of the event (e.g., executing a square root function) have in common. Clearly all events of executing the square root have certain functional properties in common (namely, they all map 4 into 2, 9 into 3, 16 into 4, and so on). But does the square root function have a temperature? Perhaps. Nothing in principle prohibits all possible instances of executing a square root from being realized in some physical form that always has a certain temperature, though that would certainly be surprising inasmuch as it is not necessary that the instances share some common physical property – and it is not hard to provide a counterexample to any claim about the temperature of a particular function just by

putting the computer in a freezer!). And even if there were such a property, it would very likely be an adventitious correlation that is highly unlikely to be of computational significance. The point is that the function is defined and has its significance in the arena of mathematical operations, not in the arena of physical processes and the two are not type-equivalent. Now what about time: Does the square root function have duration? The same answer applies. It does if every token occasion on which it occurs takes approximately the same amount of time (within some limits). It could also be that parts of the function (e.g., the basic arithmetic operations involved) have an actual duration that does not vary from token to token. But if the duration varies from occasion to occasion then only token computations of that function have some particular duration. This is what I claimed in my commentary about the duration of mental events. The response my comment evoked (“most people think that mental events have durations but Pylyshyn does not”) shows how deeply the reification of time goes. The interesting thing about this reaction is not that people think it makes sense to speak of the duration of a mental event (it does, for tokens of the event), but that the intentional error is so deep that people do not notice that any assumption is being made.

If there is a temptation to reify time in the representation of temporally extended events, the temptation to reify space is at least as strong and has immediate implications for one’s theory of mental imagery. The temptation to assume an inner display is linked to the fact that images are experienced as distributed in space (just as they are experienced as distributed in time). Because they are experienced as distributed in space we find it natural to believe that there are “places” on the image – indeed it is nearly inconceivable that an image should fail to have distinct places on it. This leads naturally to the belief that there must be a medium where “places” have a real existence. We feel there must be a form of representation underlying an image such that it makes sense to ask of some object X in the image: Where (in the image) is X? Or, How big is (the image of) X? Or, Where (in the image) are X’s parts located relative to one another? If the space in question were not real, then in what sense could we ask where a particular feature is located in relation to other features? Moreover, we have seen that a feature’s being at a particular place in relation to other features in a scene *does* have an effect on how the scene is perceived. We saw this in the examples given in Chapter 1. It is the fact that the illusory line is seen to be (or represented as being) in a certain place that creates the Pogendorff illusion, illustrated in Figure 1-5 of Chapter 1. Dan Dennett (Dennett, 1991) has written on the temptation to reify mental space and time and has pointed out, correctly in my view, that to represent a certain spatial pattern does not require that one represent all its parts as distinct individual elements situated somewhere in space. For example in representing a checkerboard one need not represent each of its squares at each possible location on the board (as assumed when one speaks of the image of the checkerboard being laid out in functional space). Dennett suggested that rather than explicitly representing each part of the imagined scene and each feature of the scene at once, the visual system may be just assigning coarse labels to regions (“more of the same here”). But this leaves a puzzle: Where is “here” and where is the coarse label placed, if not on the dubious inner display? Certainly “here” does not refer to a place on the retina or on the objects in the world (unless we are going around putting real labels on things!). The idea of labeling parts of a scene is ubiquitous and plays an important role in computational theories of vision, as we saw in Chapter 3 (especially in 3.1.1.1). But what is it that is being labeled? Can we label part of a representation or part of an experience? Many psychological theories of pattern detection (Ullman, 1984) and of visual attention (Watson & Humphreys, 1997; Yantis & Jones, 1991) speak of placing “tags” on certain stimulus objects (for example, to tell us that the item has already been processed, as in counting). If the tags must be located at a precise location in the representation, then such theories must tacitly assume an inner display of some sort – or at least a representation in which it makes sense to refer to distinct *locations*. Is there any sense that can be

made of talk of labeling or of tagging places in a scene that does not assume an inner image that reifies space?

Space has been subject to a great deal of analysis long before the debate over the nature of mental images surfaced in cognitive science. Space has always been viewed as a sort of receptacle that contains matter without itself partaking in the laws of nature – a view that had to be abandoned with the acceptance of general relativity and the notion of non-Euclidean or curved space. Even with this change in status, space (along with time) retained its special place among physical properties. Emmanuel Kant claimed that our intuitions of space and time were prior to our concepts of properties – that they were *a priori* categories of mind, presumably preconceptual and innate. The French mathematician, Henri Poincaré (Poincaré, 1963), taking a more empirical stance, wondered (and provided a partial answer to) the question how we could ever learn that space has three dimensions, given our multidimensional and multimodal experience of it. And more recently a brilliant young French graduate student named Jean Nicod, wrote a dissertation speculating on how we could construct a three-dimensional world from the sensory experiences available to us (Nicod, 1970). He put the problem differently. He took space itself to be a set of formal properties defined by the axioms of Euclidean geometry, and asked how the mind could develop a model of these axioms based on sensory experiences, together with a small number of basic perceptual categories (primarily the notion of volumetric containment). Nicod's starting point was the one we must take seriously. To say that something is spatial is to claim that it instantiates (or is a model of) the Euclidean axioms of plane geometry, at least in some approximate way (perhaps the model might be what is sometimes referred to as "locally Euclidean"). In addition, in speaking of something being spatial one usually implies that it is also metrical (or quantitative), which means that certain of its properties also fall under the metric axioms¹⁴ (see, for example, Luce, D'Zmura, Hoffman, Iverson, & Romney, 1995; Palmer, 1978). Thus a rigorous way to think of the claim that images are spatial is as the claim that certain properties of images (not of *what* they represent, but of their inherent *form* or their material instantiation) can be mapped onto, or fall under, certain formal mathematical axioms.

The claim that mental images have spatial properties in this formal sense is clearly much stronger than the empirical facts warrant. Indeed, even visual space fails to meet many Euclidean requirements (Attneave, 1954; Luce et al., 1995; Suppes, 1995; Todd, Tittle, & Norman, 1995). What exactly one means by images having spatial properties is thus not very clear. Nonetheless there are some properties of imagery that most writers have in mind when they claim that images are spatial. Before I examine the possibility that images may actually be displayed on a real physical surface in the brain, I should stop to ask what reasons there are (other than the pervasive temptation to reify mental qualities discussed above) for thinking that some form (physical or other) of space may be involved in representing spatial properties in mental imagery. There are obviously many reasons for thinking that mental images at least *encode* metrical spatial properties in some way. For example, distances on an image appear to map onto distance on the represented world in a systematic (and monotonic) way: people are able to make judgments of relative size and even some (limited) absolute judgments of size from their experienced images (Miller, 1956; Shiffrin & Nosofsky, 1994). Even if one agrees with the analysis I provided in Chapter 6, attributing the scanning effect to a mental simulation of what would happen in the real visual situation, this still requires that relative distances be encoded and used in the simulation. The ability to represent relative spatial magnitudes is surely a necessary part of our ability to recall and reason about geometrical shapes, since shape is defined in terms of relative spatial dimensions. Consequently it is reasonable to think that we use some encoding of spatial magnitudes when think using mental images.

Spatial intuitions are very powerful and appear, at least *prima facie*, to be operative when we reason about

¹⁴ In most formulations there are 4 metric axioms: If d is a metric (or distance) measure then: (1) $d(x,y) = 0$; (2) If $d(x,y)=0$ then $x = y$; (3) $d(x,y) = d(y,x)$; and (4) $d(x,z) = d(x,y)+d(y,z)$.

space while relying on our mental images, just as they are when we reason with the aid of a diagram (a topic I pursue in chapter 8, section 8.3). If we are to shun the intuitively satisfying notion that we do this by drawing an inner picture in our mind, we will need some account of how we solve such problems at all. Nobody is yet in a position to provide a detailed account of this sort. On the other hand, neither are the proponents of the picture-theory: Even if there were real 2D diagrams in the head we have no idea how they could be used. What we do know is that sooner or later the pictorial information has to be interpreted and transformed into a form that can enter into inferences, i.e., a form that meets the requirements of compositionality and productivity discussed in Chapter 8, section 8.2 (and taken up in detail in Fodor, 1975; Fodor & Pylyshyn, 1988).

The spatiality of images arises not only from the fact that we can retain distances and shapes in memory (often in visual memory). There are other properties of images that give them their apparent spatiality. Among these are the experiments discussed earlier (e.g., mental scanning), the more robust of which are those in which people “project” their image onto the world they are perceiving. They are discussed in the next section, where I suggest that some spatial properties inferred from imagery experiments are real, but do not arise from the spatial nature of the images themselves. These spatial properties, I argue, can be attributed to the real spatial nature of the sensory world onto which they are “projected.” Another source of evidence for the spatial nature of images comes from the way that images connect with the visuomotor system. These two sources of evidence for the spatiality of images are discussed in the following sections.

7.3 Inheritance of spatial properties of images from perceived space

In many imagery studies subjects are asked to imagine something while looking at a scene, thus at least phenomenologically superimposing or projecting an image onto the perceived world. Although it has been amply demonstrated (see the review in O'Regan & Lévy-Schoen, 1983) that true superposition of visual percepts does not occur when brief visual displays are presented in sequence, or across saccades, the impression that superposition occurs in cases of imagery remains strong. What happens, then, when a mental image (whether constructed or derived from memory) is superimposed on 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 guided by what is experienced as the projected image. A simpler way of putting this is that one simply *thinks* of imagined objects as being located in the same places as certain perceived ones. For example, in the (Podgorny & Shepard, 1978) study the investigators asked subjects to imagine a letter-like figure projected onto a grid, such as shown in Figure 7-1) and to indicate as quickly as possible whether a spot that was then displayed in the grid was located on or beside the imagined figure. In the visual version of the task, subjects were faster when the spot was on the figure than when it was immediately beside it, faster when it was at certain stroke intersections (such as in the corners or the T junction) than when it was in the middle of a row or column, and so on. When subjects were asked to imagine the figure projected onto the empty grid, the pattern of reaction times obtained was very similar to the one obtained from the corresponding real display. This result was taken as suggesting that the visual system is involved in both the visual and the imaginal cases. But a more parsimonious account is that in “imagining” the figure in this task, observers merely attended to the rows and columns in which the imagined figure would have appeared. We know that people can indeed direct their attention, or assigning an index, to each of several objects in a display or to conform their attention to a particular shape. As I argued in chapter 5, either focusing attention in this way or assigning several indexes to certain objects, 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 regions was more effective in enhancing detection in those regions than the instruction to superimpose an image over the region.

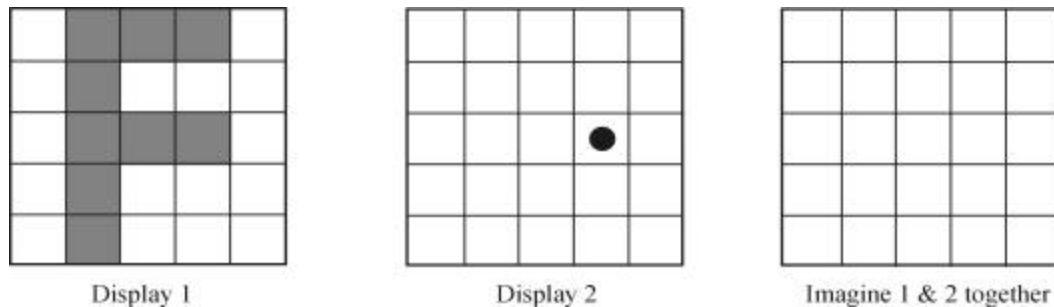


Figure 7-1. Observers were shown a figure (display 1) which they then had to retain as an image and to indicate whether the dot (display 2) occurred on or off the imagined figure. The pattern of reaction times was found to be similar to that observed when the figure was actually present. (based on Podgorny & Shepard, 1978).

A similar analysis applies in the case of other tasks that involve responding to image properties when images are superimposed over a perceived scene. If, for example, you imagine the map used to study mental scanning (discussed in Chapter 6) superimposed over one of the walls in the room you are in, you can use the visual features of the wall to anchor various objects in the imagined map. In this case, the increase in time it takes to access information from loci that are further apart is easily explained since the “images” or, more neutrally “thoughts,” of these objects *are actually* located further apart. What is special about such superposition cases is this: The world being viewed contains rigid 2D surfaces which embody the properties expressed by spatial axioms – i.e., Euclidean and other mathematical properties of metrical space literally apply to it – so the combined image/perception inherits these properties. How does this happen? How does cognition obtain access to these spatial properties and apply them to the superimposed image?

What makes it possible for certain spatial properties to be inherited from a real scene is the visual indexing mechanism discussed in chapter 5. Using this mechanism, a person can pick out a small number of objects in a scene and associated mental contents (or thoughts) with them. So, for example, the person could think that particular token objects in a scene correspond to certain imagined objects in memory. As we saw in Chapter 5 we can have thoughts such as “assume that this <object-token-1> is the beach, and that <object-token-2> is the lighthouse,” where the italicized terms are pointers to visual objects that are picked out and referenced using visual indexes. We have already seen that such visual indexes allow thoughts about certain token properties to be connected to particular objects in a scene, the way that demonstrative terms like “this” or “that” do in natural language. Thus we can literally think, “*this* is the beach,” “*this* is the lighthouse,” and so on. Given this capability, now something interesting happens. Because the scene obeys physical laws and geometrical axioms, the spatial relations among indexed objects systematically maintain their spatial relations, providing only that visual indexes remain bound to their scene objects. For example, if you indexed three points that just happened to be collinear, then regardless of what you had encoded about their relationship, they would always have an ordinal position along the line; the second point would have the relation “between” to the other two, and so on. So if you subsequently *noticed* that they were collinear or that the “between”

relation held among the three points, you would be guaranteed that it was consistent with the axioms of geometry and with their being on a rigid surface, whether or not you remembered or noticed that the points you had picked out had other properties and whether or not you knew that they maintained their fixed locations while you examined other properties of the scene. Such a guarantee would be underwritten, not by *your knowledge* of geometry (and, in particular, not by your knowledge of the formal properties of the relation “between”), but by the physical facts about the world you are looking at and the fact that your visual system is able to detect certain relations that hold among objects in that world. In other words, by picking out (i.e., indexing) certain objects in the world, and by binding certain thoughts to these objects, you are able to draw conclusions about their configuration by *visually* noticing these properties. Without such a display you would have to draw inferences based on the postulates of geometry.

This point connects directly with several of the phenomena that have been cited in support of the claim that mental images are spatial. For example, if you could individuate and refer to several objects in a visual scene, you could then associate certain objects or properties in memory with each of these perceived objects. Then you could literally scan your attention (or even your gaze) from one to another of these places. Many other imagery phenomena might also be attributable to real vision operating over a display. For example, with real vision taking part in the imaging process one could deploy attention in a way that could produce certain visual illusions (as suggested in section 6.5.2). Such attentional phenomena are different from a real superimposition of an image over the scene. For example, in the scanning case all you need to be able to do is recall (from *some* representation of the scene) where point B was in relation to point A. The effect does not depend on a detailed picture being projected, only on attention and visual indexes being allocated in certain ways. By anchoring a small number of imagined objects to real objects in the world, the imaginal world inherits much of the geometry of the real world.

7.3.1 Scanning when no surface is visible

What if the scanning experiment were to be carried out in the dark so that there were no visible objects to put into correspondence with the objects in your memory? One prediction would be that you would have trouble with this task, since it is known that after a few minutes of looking at such a display (called a Ganzfeld), vision becomes very unstable and it even becomes difficult to keep track of where you are looking in relation to where you were looking earlier (Avant, 1965). Such featureless displays are very rare – almost any surface has texture features, as can easily be seen, for example, in first and second derivative images which show luminance discontinuities (e.g., in Marr, 1982). But what if you generate a new image of something without at the same time viewing a real scene? What if you imagine a scene with your eyes closed? The first thing that can be said is that the results in such cases are much less robust and more easily influenced by task strategies and beliefs. It is not even clear that mental scanning experiments can be carried out in total darkness (for more than a very short time) since even viewing a single light in total darkness leads to illusory motion called the autokinetic effect (in which a stationary light appears to move around autonomously). As argued earlier (in section 6.5.3) there is some reason to believe that smooth imaginal scanning, in which the focus of the scan passes through intermediate points, is not possible under these conditions. Notwithstanding these problems, it may still be possible to generate the relevant time-distance relationships without having a spatially laid out scene either in the visible world or in your head. All you need is a way to represent distances. While there is no general theory of how such magnitudes are encoded, we know that they are encoded, not only in humans but also in animals (Gallistel, 1990). Notice that being able to encode magnitudes is quite different from the assumption of a spatial image, since analogue encoding

schemes lack the multi-dimensionality and other properties of images¹⁵.

There are few published studies of mental scanning that did not involve a visual display. In one of these (Pinker et al., 1984), discussed briefly in section 6.3.2.1, locations of three objects in a 2D matrix were described verbally, along with the location and direction of an arrow. The task was to judge whether the arrow pointed to one of the objects and the data showed that it took longer to make this judgment when the points were further away from the location of the arrow. Why should it take longer to judge whether an arrow is pointing at an object in memory when the object is imagined to be further away? The Pinker et al. study appears to rule out an explanation based on inheriting spatial properties from a visual display and suggests that it might be possible to get the same effect by appealing to perceived space in modality other than vision – in fact it may be possible to exploit a more general “spatial sense” which may involve proprioceptively sensed space or the space of potential motor movements. We explore this possibility in the next section.

7.3.2 The exploitation of proprioceptive or motor space

Where does the strong impression of the spatial nature of our images come from? What, other than the subjective experience of seeing our mental image laid out in space, prompts us to think that images are in some important sense spatial? One simple answer is that when we imagine a spatially distributed layout of objects and shapes then something is indeed spatial, namely the layouts and shapes that we are imagining. It is the things that we think about, not the patterns in our brain that we think with, that are spatial. This simple answer should carry more persuasive force than it typically does: just as when we imagine a round thing, or a green thing, or a furry thing, it is the thing we are imagining that has those properties, not our image. But there is one difference in the case of spatial properties of mental images: If you close your eyes and imagine a familiar spatial layout (say a map of your city), you feel the image is “out there in front of you” in the sense that you can actually point to parts of it. But what are you actually pointing at when you do this? I have already suggested that when you do this while looking at a scene (“projecting” your image onto the visible world) you may be binding objects in your image to objects in the scene, thus providing the imagined objects with an actual location in the world and thereby inheriting the spatial properties of the real external space. Might it be that when you do this with eyes shut (or in the dark) you bind objects in your image to places in a *proprioceptively sensed* space? If that were so then there would once again be a real space over which scanning could occur, but as in the case of projecting images onto a perceived scene, the space would not be in the head; it would be real space in the world or in the body, sensed through proprioceptive, haptic, kinesthetic, auditory, or any other perceptual modality that was operating when your eyes were closed.

Our sense of space is extremely well developed and is used not only for thinking of concrete spatial patterns, but also for individuating and thinking of abstract ideas. In fact, sign languages (like American Sign Language) make use of locations in speaker-hearers’ perceived space to individuate ideas and refer back to them. An entire topic can be located at a place in allocentric space and then when the speaker wishes to refer to that topic he or she simply points to the empty space where the topic had been “deposited”. You can easily demonstrate for yourself how good we are at localizing things in the space around us. Close your eyes and point to things in the room around you; you may find that you can point accurately and without hesitation – even to things behind you. Given this skill it seems plausible that we can utilize it for binding objects in our image to places in our immediate proprioceptive space, even in the absence of vision. What this would give you is the sensed spatial quality of images without requiring the images to *have* space themselves. The

¹⁵ Note that I have no quarrel with those who propose analogue representations of magnitudes. This is a very different matter from the assumption that there is an analogue for an entire system of magnitudes corresponding to space. The trouble is that unlike the hydraulic analogue of electric flow, an analogue of space would have to encompass many properties (and at least 4 dimensions) and meet a large number of constraints, such as those embodied in Euclidean axioms. It would have to exhibit such general properties as Pythagoras’ Theorem, the single point-of-view requirement discussed in (Pinker & Finke, 1980) and countless other properties of projective geometry (including the metrical axioms, see section 7.1). It is no wonder that nobody has been able to think of an analogue to space-time other than space-time itself.

difference between appealing to images (even proprioceptive images) and appealing to what I have been calling a *spatial sense*, is that the latter does not assume any particular form of representation, only the skill to pick out locations that are currently being sensed, through whatever ambient inputs that are available for use to calibrate locations.

Studies of the recall of locations have demonstrated a surprisingly facile skill that allows us to sense and to recall where things are around us. (Attneave & Farrar, 1977; Attneave & Pierce, 1978) showed that the ability to recall locations was extremely good and the accuracy was about the same whether the locations were in front of or behind the observer. When observers looked at a row of 7 objects located in front of them and were later seated with their backs to where the objects had been, they could recall their relative locations almost as well as they could if they were asked to imagine them as being in front of them. It thus appears that observers could easily take one or another of two perspectives 180 degrees apart in recalling the locations of objects in a room. They could do this even without generating a detailed visual image of the objects. (Attneave & Farrar, 1977) reported that when subjects were asked about the relative location of two objects "...the second object was located in space (subjects typically said) and the question was answered accordingly, before the object was clearly pictured. This is logically consistent with the report that images were evoked or constructed, rather than continuously present: one must decide *where* to draw the picture before drawing it." It seems that deciding where the item is occurs primitively, quickly and accurately. Attneave & Farrar also report that when an image of the scene behind the subject's head did occur it was not accompanied by a sense of turning-and-looking. They remark, "This contradicts the view that a visual image must correspond to some possible visual input. ... We are being told, in effect, that the *mind's eye* has a cycloramic, 360 degree field." This suggests that representing location-in-space is rather different, both phenomenologically and behaviorally, from what we usually consider to be visual images (which have a "visual angle" about the size of real vision Kosslyn, 1978).

The study of the "sense of space" (a term chosen so as not to prejudge the question of whether the space is visual, proprioceptive, motoric, or completely amodal) has only recently been conducted under controlled conditions and has led to some unexpected findings. For example, it was shown that although people have a robust sense of the space around them, this space tends to be calibrated primarily with respect to where the person is located at any moment. The reference point, where people perceive themselves as being, also depends on their motor activity; as they move about they automatically recalibrate their proprioceptive sense of space, even if they move about without benefit of vision. (Rieser, Guth, & Hill, 1986) had subjects view a set of target objects (which also served as alternative viewing points) and then had them either walk blindfolded to specified points or merely imagine walking to these points, and then to indicate (by pointing) the direction of other targets. They found that people were faster and more accurate when they had moved on their own volition to the designated locations, even though they did not have the benefit of visual information in doing so. This was also true for pure rotation. People who had rotated their body appeared to instantaneously recalibrate their orientation so they could point to the location of various targets, whereas

when they merely imagined themselves to rotate, their localization of targets was slow and deliberate¹⁶ (Farrell & Robertson, 1998; Rieser, 1989).

In his commentary on my paper on mental imagery, David Ingle (Ingle, 2002) describes his own case of long-term visual persistence (which he views as an extreme type of mental image). He observed that the location of his image remains fixed in allocentric space as he moves or turns around. But if he observes an object held in his hand, and closes his eyes, his image of the object moves when he moves his hand. This sort of anchoring of perceived location of an image to proprioceptively perceived locations is consistent with the notion that proprioceptive cues are used to locate images in real space. We will see another example of this principal in the next section when I discuss the finding that a visually guided motor skill (smooth pursuit), which responds only to real perceived motion and not imagined motion, also responds to proprioceptively perceived motion.

The idea that the spatial quality of a mental image derives from its connection with proprioceptive and motor systems is further supported by the work of Ronald Finke (Finke, 1979) on the interaction of mental images and visual-motor control, and by findings regarding the role of motor activity in relation to locations of objects in mental images. The work was carried out to support a version of the picture theory of mental imagery, but as we will see, it does not require the involvement of anything pictorial; only the position of imagined objects in relation to the environment is relevant.

Consider the SR compatibility findings of (Tlauka & McKenna, 1998), which showed that when you respond using crossed and uncrossed hands, it takes longer to respond to features in the opposite side of your mental image. From this (Tlauka & McKenna, 1998) concluded that stimulus-response compatibility factors affect reactions to locations in images just as they do in real displays. But what the result actually shows is that in orienting to objects imagined to be in certain locations in (real) space, observers orient to these same locations just the way they would have if the objects actually were at those locations. Thus it is not surprising that the same phenomena are observed when observers react to objects in real displays as they do to objects merely imagined to be located “out there” in these same locations. We will have a closer look at this phenomenon again below.

The (Finke, 1979) studies are striking because they involve a detailed analysis of the interaction of imagery, visual-motor coordination, and “felt location” of objects. In a series of ingenious experiments, Finke showed that the well-known adaptation to displacing prisms could be obtained using imagery alone. In the original visual adaptation phenomenon observers wore prism goggles that displaced the apparent location of everything in view by some fixed angle (e.g., 23 degrees). After wearing the goggles for some period of time, observers became adept at reaching for things (and also walking around). When the goggles are later removed, observers had the opposite experience – objects looked to be mislocated in the opposite direction to

¹⁶ It is possible to make too much of the involvement of motor activity in mental imagery. I have suggested that we use our proprioceptive/motor space in deriving the spatial character of mental images. But others have suggested that the motor system is involved in transforming mental images as well. The role that motor mechanisms play in the transformation of visual images is far from clear, notwithstanding evidence of correlations between visual image transformations and activity in parts of the motor cortex (Cohen et al., 1996; Richter et al., 2000), or the influence of real motor actions on visual transformations (Wexler, Kosslyn, & Berthoz, 1998). Some of the cortical activity observed during both motor performance and the mental transformation of visual images, may reflect the fact that these areas (e.g., posterior parietal cortex) compute some of the higher-level functions required for extrapolating trajectories, for tracking, for planning, and for visuomotor coordination (Anderson, Snyder, Bradley, & Xing, 1997). Since many of these functions also have to be computed in the course of anticipating movements visually, it is reasonable that the same areas might be active in both cases. While studying the interaction of imagery and the motor system is clearly important, at the present time we are far from justified in concluding that dynamic visual imagery is carried out *by means of* the motor system (or that visual operations *exploit* motor control mechanisms). This way of speaking suggests that our motor system can grasp and manipulate our images, a view that unfortunately reinforces the general tendency to reify the world that pervades much mental imagery theorizing.

the way the prisms had shifted them in the adaptation phase. Finke's studies are of special interest to the present discussion because they illustrate the way in which projected images can work like real percepts in certain ways – in particular in respect to perceptual-motor coordination. They also illustrate an important role played by visual indexes in accounting for certain results in studies on mental imagery.

In one study, (Finke, 1979) asked subjects to imagine seeing their (hidden) hand as being in certain specified locations. The locations where he asked them to imagine their hand to be, corresponded to the errors of displacement actually made by another subject who had worn displacing prisms. He found both the pattern of adaptation and the pattern of after-effects exhibited by observers who only had imagined feedback to be similar to that exhibited by observers who actually wore displacing prisms. Now it is known that adaptation can occur to solely verbally presented error information (Kelso, Cook, Olson, & Epstein, 1975; Uhlarik, 1973), though in that cases (and in contrast with the case where the hand is continually viewed), 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 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 visually presented error information. When subjects did not move their arm, or if they just imagined moving their arm, the results were like those obtained when they were only given verbal feedback – i.e., slow adaptation effects that transfer to the nonadapted hand. From these results, Finke concluded that adaptation to imagined hand location taps into the visual system at the same “level” as that of visually presented information. But do these results really require that we appeal to the visual system, as understood in chapter 2, or can they be explained in terms of the orienting of attention to *real* places in a visual display?

The generally accepted view of what goes on in prism adaptation experiments is a recalibration between where subjects are looking or visually attending and either where they feel their arm is located or the motor commands they must issue in order to move their arm correctly (the so-called re-afferent signal). The exact way this happens has been the subject of some debate (Howard, 1982), but it is generally accepted that important factors include the discrepancy between the seen position and the felt position of the hand (or the discordance between visual and kinesthetic/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 ensuing adaptation. But it also appears that where the subject attends is equally important (Canon, 1970, 1971). In some cases even an immobile hand can elicit adaptation providing the subject visually attends to it (Mather & Lackner, 1981). Thus the imagery condition in Finke's study provides all that is needed for adaptation – without making any assumptions about the nature of imagery. In particular, subjects direct their gaze towards a particular (erroneous) location where they are in effect told to pretend their hand is located, thus focusing attention on the discordance between this viewed location and their kinesthetic and proprioceptive sense of the position of their arm.

Ian Howard (Howard, 1982) has provided a thorough discussion of the conditions under which one gets more or less adaptation. The most important requirement for adaptation is that the discordant information be salient for the subject, that it be attended, 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. It makes sense, therefore, that merely telling subjects where their hand is would 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.

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 study of S-R compatibility. None of these results require that imagery feed into the visuomotor system. Indeed, these two cases involve *actual* visual perception of location (i.e., there really are some visible features located in the relevant locations). The only information that needs to be provided by the mental image in order for adaptation (as well as the S-R compatibility effects) to occur, is information about the location of some indexable visible features in the scene, where the hand can be imagined to be located (by binding the idea of the hand to a visible feature through the index). It does not require that the image provide shape, size, orientation, color, *or any other visual information*, other than location where things are imagined to be.

Rather than support the motion that imagery feeds into the visuomotor system at the same level as vision, the evidence we have considered provides support for the notion that vision and motor control are closely connected precisely because the spatiality of images derives directly from the spatiality of the real world that is provided by proprioceptive and kinesthetic perception. This connection of image locations with both proprioceptively sensed space and with potential actions of the motor system is much more important than has generally been recognized. It extends the earlier argument about imaginal space being inherited from vision, to the proposal that imaged space can be inherited from a wider range of space-sensing modalities. Just as we can scan our eyes or our attention in relation to the visual perceived space when we examine images that are projected into a visual scene, so we can move our hand from one imagined place to another, and when we do so we demonstrate the spatiality of images.¹⁷

7.4 The search for a real spatial display

There are many reasons for resisting the conclusion that images are themselves spatial. I have presented a number of arguments suggesting that many of the experimental findings that led people to claim that images are spatial (or at least that they have “metrical” properties) can be explained more parsimoniously in terms of the use of tacit knowledge to simulate what would happen if observers actually saw the appropriate event actually taking place (e.g., if they were to actually examine a real map). I have also presented some arguments suggesting that the spatial character of images may derive from the way that imagined objects are attached (using FINST indexes) to perceived features in the world. When a visual surface is not available, sensed locations in proprioceptive space might be used instead. What will not do, as an explanation, is to appeal to a “functional space” since the only way such an appeal can be explanatory is if the functional space is taken as a simulation of real space. The one alternative that remains viable is the claim that images appear spatial because they are actually realized in the brain on real 2D spatial displays. If we are convinced that images are different from other forms of thought and if we think that part of their difference is that they are laid out in space (or that they are “depictive”), then a reasonable strategy might be to take the bull by the horns and try to find an actual literal spatial medium in the brain. This approach at least has the virtue of making an explicit testable claim, which, despite its initial implausibility, would provide some explanatory advantage over an illusive metaphor or an appeal to the misleading notion of a “functional space.” Despite the

¹⁷ In chapter 5 I discussed a theoretical mechanism for binding objects of thought to perceived objects in a scene. This mechanism is the FINST visual index. In the original paper where this mechanism was introduced, I proposed a parallel mechanism that binds objects of thought (or of visual perception) to proprioceptively sensed objects or locations. These indexes were called *Anchors* and they play an important role in perceptual-motor coordination along with FINST indexes.

advantage of this approach, surprisingly few picture theorists have been willing to endorse the assumption that there is a literal spatial display in the brain. In recent years, however, much of the work directed at supporting the picture-theory of mental imagery has been carried out within neuroscience, and much of it has involved the search for a spatial display (using new techniques such as neural imaging). The hope has been to find a 2D display along the lines of the CRT metaphor introduced several decades ago (Kosslyn et al., 1979). In what follows I will examine the recent neuroscience work from several perspectives, beginning with a general methodological discussion of the status of neuroscience evidence in the imagery debate.

7.4.1 Aside: Does biological evidence have a privileged status in this argument?

A number of writers appear to feel that neuroscience evidence renders all previous behavioral evidence obsolete. Stephen Kosslyn himself laments the indecisive nature of mental imagery research of the past 25 years and speaks of the new neuroscience-based research as finally being able to provide a clear and decisive answer to the nature of mental imagery (Kosslyn, 1994). 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 has cast doubt on the picture theory. Even if we found real colored stereo pictures displayed on the visual cortex, the problems raised thus far in this and the previous chapter would remain and would continue to stand as evidence that these 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* might be displayed on the cortex it could not be what is responsible for the patterns of reaction times observed in the scanning studies because, as I argued in section 6.3.1 and 6.3.2, those patterns do not reflect properties of mental architecture, but properties of what subjects know. Similarly the question of what is responsible for the facilitation of recall and problem solving that accompanies the phenomenology of mental imagery requires a psychological process theory to link any proposals about the nature of mental images with actual performance data. It is important to understand that the mere fact that the data are biological does not give it a privileged status in deciding the truth of a psychological process theory, especially one whose conceptual foundations are already shaky.

In examining the behavioral evidence so far I have distinguished two distinct types of claims about the representations underlying mental images. The first concerns the nature or the format of mental images and the second concerns the nature of the mechanism used in processing them. We saw that although these may be related questions, they are also largely independent, since it is logically possible for the visual system to be involved in both vision and mental imagery and yet in neither case use picture-like representations. Similarly it is possible for representations to be topographically organized 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 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 constituents and the rules by which the complex is put together (the way the meaning of sentences is compositional and depends on the meaning of its constituents together with the way they are syntactically put together). But the physical instantiation of any representation that meets the requirement of compositionality will itself be compositional (Pylyshyn, 1984a, pp 54-69; 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

6.4.1 are met by any compositional physical symbol system, not just a depictive one. Note that in a digital computer, representations are both 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 (especially when they use bitmapped codes, such as GIF or JPG or BMP), their topographical distribution does not mirror the physical layout of the picture. Thus 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. In the present state of neuroscience, it remains highly unclear how information processing mechanisms, representations and other theoretical entities map onto brain structures, and consequently it is unclear how such evidence can address the question of the format of thought, including the format of mental images. In what follows I will look at some of the evidence as it applies to the study of mental imagery. In the course of this review it will become clear that what neuroscience evidence is being directed towards is what it is best at; locating active regions in the brain. As a result the technique is tailor-made for testing the literal picture-in-the-head theory of mental imagery.

As we saw earlier, the question of whether mental imagery uses the visual system is intimately tied to the question of what constitutes the uniquely visual system. If the question is merely about whether some mechanisms used in vision are also used in visual imagery, then the answer is clearly yes, though for the uninteresting reason that they both involve accessing memory and making inferences. 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 early 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 examining the neuroscience evidence concerning which areas of the brain are involved in mental imagery and in visual perception. It is to this question that I now turn, beginning with an examination of the neural activity evidence cited in support of the claim that mental images are realized in a topographic or spatial display in the brain, and then considering some of the clinical evidence from brain damage.

7.4.2 The argument from differential activity in the brain

An argument along the following lines has been made in the recent neuroscience literature (Kosslyn, Pascual-Leone et al., 1999; Kosslyn, Thompson, Kim, & Alpert, 1995). Primary visual cortex (Area 17) is known to be organized retinotopically (at least in monkey brain). So if the retinotopic visual area were active

when subjects generate mental images¹⁸, it would suggest that (1) the early visual system is involved in some aspect of processing visual mental images, and (2) during imagery the visual system receives inputs in the form of a retinotopic display. In other words during imagery the cognitive system generates a display that is laid out in a spatial or “depictive” form (i.e., like a two-dimensional picture) in primary visual cortex, and this display is then interpreted by early vision.

Writers espousing the picture theory routinely cite evidence showing that the early vision area of the brain is organized retinotopically. For example, one of the most widely cited papers is a study by (Tootell, Silverman, Switkes, & de Valois, 1982), which shows that there is a pattern of activation in monkey visual cortex that closely resembles a pattern of lights that the monkey viewed. Tootell et al. trained macaques to stare at the center of a pattern of flashing lights, while the monkeys were injected with radioactively tagged 2-deoxydeoxyglucose (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 (except for a cortical magnification distortion). 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, as shown in Figure 7-2. This led many people to conclude that a picture, corresponding to what we see, appears in primary visual cortex during visual perception. Although no such maps have been found during 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.

¹⁸ At present one of the primary sources of experimental neuroscience evidence on brain activity during mental imagery comes from neural imaging (particularly fMRI). This fairly recent technique, which is extremely important in clinical neurology, relies on assumptions about the relation between blood flow and metabolism, and between metabolism and cognitive processing in the brain. Yet it is far from clear what kind of neural activity is indicated by increased blood flow; whether, for example, it indicates activation or inhibition (e.g., the active attempt to suppress otherwise disrupting visual activity); whether it is associated with the same activity that is responsible for the sorts of behavioral measures discussed earlier (e.g., reaction time functions), whether it related to the experience of “seeing,” or something entirely different. It is even problematic whether increased cerebral blood flow indicates increased information processing activity (Fidelman, 1994; Haier et al., 1988). There is also concern that the widely used subtractive technique (in which the activity map associated with a control condition is subtracted from the activity map associated with an experimental condition) has problems of its own (Sergent, 1994), as well as being predicated on the assumption that the function under study is a modular one that involves activity in a particular brain region each time it occurs, regardless of the circumstances (Sarter, Berntson, & Cacioppo, 1996). Even processes known to be functionally modular, such as the syntactic aspect of language processing, often give inconsistent neural imaging results (Démonet, Wise, & Frackowiak, 1993), so we ought to be wary when drawing conclusions about functions like mental imagery, which the empirical data give us every reason to believe is non-modular. Problems such as these may eventually be resolved by better technologies for more precise localization and measurement of brain activity in time and space. But the value of neuroscience data will increase even more when we develop a better understanding of the questions that we need to address. For the time being we should treat neural imaging evidence the way we treat all other essentially correlational evidence, such as reaction time, error rate, or ERP – as indirect indexes whose validity needs to be independently established, rather than as direct measures of the variables of primary interest. Even the use of the intrusive method of transcranial magnetic stimulation (rTMS), which gets around the problem of purely correlational evidence by effectively disabling an area of the cortex, has its own problems, and cannot be assumed to provide direct evidence of causal mechanisms. For example, in a recent paper (Kosslyn, Pascual-Leone et al., 1999) showed that if area 17 is temporarily impaired using rTMS, performance on an imagery task is adversely affected (relative to the condition when they do not receive rTMS), which Kosslyn interpreted as showing that the activation of area 17 plays a causal role in imagery. 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 similar tasks *not* involving imagery.

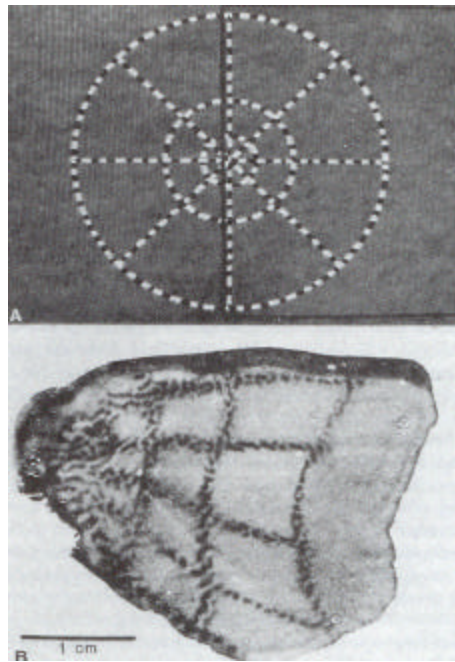


Figure 7-2. The top photograph (A) shows the stimulus used by (Tootell et al., 1982) to demonstrate the retinotopic organization of the visual cortex in the macaque. The monkey stared at the top pattern, formed by flashing lights, while it was injected with a radioactively tagged tracer. Figure B shows the pattern of radioactivity activity recorded on the Macaque's primary visual cortex.

The idea that in mental imagery, cognition provides input into the early visual system which is then “perceived” and given a (possibly new) interpretation, is very much in keeping with the views developed earlier from behavioral evidence. It is also in keeping with the subjectively satisfying picture theory of mental imagery. Given this background, it is no surprise that those who hold the picture theory would receive any evidence of the involvement of early vision in mental imagery with a great deal of enthusiasm. Evidence such as that of Tootell et al. has been hailed as the critical evidence that shows how vision and imagery are related: viz., they both involve generating an image in primary visual cortex which is processed by the visual system. We have already seen some reasons to doubt the picture theory, so one might perhaps wonder what the neuroscience results tell us about either vision or imagery. The relevance of the Tootell et al. results to the picture theory of mental imagery will be discussed in section 7.5 (especially 7.5.4).

With this sketch as background, we can now ask whether there is indeed evidence for the critical involvement of early vision in mental imagery and, if so, what it means. There appears to be some evidence that mental imagery involves activity in areas of striate cortex associated with vision, though the question of whether it is necessary or sufficient for mental imagery is far from being univocally accepted. Most of this evidence has come from studies using neural imaging to monitor regional cerebral blood flow, or from studies of brain-damaged patients. While some neural imaging studies report activity in topographically organized cortical areas (Kosslyn, Pascual-Leone et al., 1999; 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 et al., 1997; Fletcher, Shallice, Frith,

Frackowiak, & Dolan, 1996; Goldenberg, Mullbacher, & Nowak, 1995; Howard et al., 1998; Mellet, Petit, Mazoyer, Denis, & Tzourio, 1998; Mellet et al., 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). 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).

In order to support the “cortical display” version of the picture theory it is important not only to show that the visual system is involved in imagery, but also that the areas involved are the topographically mapped areas of cortex and that their involvement is of the right kind. In particular it is important that the topographic organization reflect the spatial properties of the phenomenal image. Very few neuroscience studies meet this criterion, even when they show that the visual areas are activated during mental imagery. One of the few examples of a finding that has been assumed to meet this criterion was reported in (Kosslyn et al., 1995). This paper describes findings that relate a specifically spatial property of mental images (their size) to a pattern of neural activity. It 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, it has been taken to support the claim that the activation of visual cortical areas during mental imagery corresponds to the activation of a cortical display which maps represented space onto cortical space. It is for this reason that (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 this conclusion is premature since the spatial distribution of brain activity is not the kind that would explain the phenomenological and behavioral findings concerning image size and image distances. Even if the cortical activity that shows up in the PET scans corresponds to a mental image, the evidence only shows that a mental image experienced as being larger involves activity that is located in areas where larger retinal images would project. But in the case of vision, the reason that larger retinal images activate the regions that they do is related to the way that the visual pathway projects from the periphery of the retina to the occipital cortex. Thus the PET result is not the same as finding that a pattern of activation maps the size of a mental image onto a metrical spatial property of its cortical representation (Fox et al., 1986). In particular, the PET data do not show that image size is mapped onto some function of the size of the active cortical region. On reflection it might not be surprising that properties such as image size do not map monotonically onto the size of the activated region, yet this is exactly what would have to happen if the activity pattern is to serve as the basis for explaining such phenomena as mental scanning. While picture theorists may not realize it, a literal metrical mapping is required by the cortical display view. The mapping does not have to be linear, but it does have to be a continuous mapping (up to the grain of neuronal cells) that preserves local topology – such as a mapping known as a homeomorphism or a locally affine transformation. This is what it means to claim (as Kosslyn does in Kosslyn et al., 1978) that images “preserve metrical information”.

A cortical display that preserves at least a monotonic function of magnitudes would also be required if the display were 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 than in a small one is that it is easier for the mind’s eye to “see” a feature in large cortical display. One could not attribute that result to the fact that some of the mental 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 findings regarding the spatial nature of mental images discussed earlier (e.g., the mental scanning result). Consequently the PET data cannot be interpreted as supporting the picture theory of mental imagery, nor does it in any way help to make the case for a cortical display theory of mental image representation. Those of us who eschew dualism are perfectly prepared to accept that *something* different happens in the brain when a different phenomenal experience occurs, consequently we take it for granted that something different must occur in the brain when a larger image is experienced. The point has never been to dispute the view that psychological properties are supervenient on physical properties, 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 on information access, and other metrical effects) and perhaps even the subjective content of mental images. If it had turned out that a larger phenomenal image was accompanied by a larger area of cortical activity (or if greater phenomenal distances were mapped onto monotonically greater cortical distances) it might have left room for a possible cortical display account of some of the classical imagery findings (though it would still have left many unanswered questions), but a merely different locus of brain activity is no help in explaining metrical effects.

7.4.3 The argument from clinical cases of brain damage

Another source of evidence that has drawn the interest of picture-theorists is evidence collected from brain-damaged patients.¹⁹ If the specifically visual areas of the brain are responsible for mental imagery, then it follows that damage to those areas should impair both vision and imagery functions in similar ways. In particular, if the damage was to a cortical display one should find very similar patterns of deficits in both vision and imagery. A number of cases of parallel deficits in imagery and vision have indeed been reported. For example, (Farah, Soso, & Dasheiff, 1992) reported that a patient who developed tunnel vision after unilateral occipital lobectomy also developed a reduction in the maximum size of her images (as determined by asking how close an image of a familiar object, such as a chair or an automobile, could be before it overflowed the edge of her image). If the cortical display were involved in both vision and imagery, and if the peripheral parts of the display were damaged, then it might explain the parallel deficits. Although it is certainly possible that tunnel vision and tunnel imagery could have a common underlying neural basis, the Farah et al. finding 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. In fact in this case there is a possible explanation for the reported finding that need not even involve the visual system, let alone a cortical screen. If one accepts the proposal that many imagery phenomena, such as those associated with different sizes of mental images, generally arise from the implicit task requirement of simulating aspects of what things would look like, then it is possible that the patient was merely reporting how a visual stimulus looked to her. In this study the patient had nearly a year of post-surgery recovery time before the imagery testing took place. During this time she would have become familiar with how things looked to her now, and was therefore in a position to simulate her visual experience by producing the relevant phenomena when asked to image certain things (e.g., to answer appropriately when asked at what distance an image of some familiar object would overflow her image). As I have often pointed out, this would not be a case of the patient being disingenuous or being influenced by the experimenter, which Farah et al. were at pains to deny, but of the patient doing her best to carry out the required task, namely to “imagine how it would look.”

The clinical phenomenon of visual or hemispatial *neglect* has also often been cited in support of the

¹⁹ Problems comparable to those faced by neural imaging studies (see note 18) also arise with the use of data from clinical populations (Trojano & Grossi, 1994). It is hard enough to find “pure” cases in which the lesions are highly localized and their locations accurately known, but when such patients are found, the evidence is even less clear in supporting the involvement of the earliest (topographically organized) areas of visual cortex (Roland & Gulyas, 1994b).

cortical display theory. In a famous paper (Bisiach & Luzzatti, 1978) reported two patients who had the classical syndrome of visual neglect, in which they tended not to report details on one side of a scene (in this case the left side). Bisiach & Luzzatti found that both patients also tended to omit reporting details from the left side of their mental images. More interestingly, this did not appear to be a problem of memory since they could recall the details that appeared to be missing in the left side of their image if they were asked to imagine turning around and then to report their image of the scene viewed from the opposite perspective. This is indeed an interesting phenomenon and is frequently cited in favor of a cortical-screen view of mental imagery (after all, it's hard to think why a symbolic representation would favor one side of a scene over another). The explanation in terms of a cortical display appears to be straightforward: if one side of the display is damaged then one would expect both visual perception and visual imagery to be subject to neglect on the same side. However, it has since turned out that visual neglect and imaginal neglect are dissociable. There are patients who neglect one side of a visual scene but do not neglect one side of their image, patients who neglect one side of their image but do not neglect one side of a visual scene, and even patients who neglect one side of a visual scene and neglect the other side of their image (Beschin, Basso, & Sala, 2000).

Notwithstanding the problems of replicating the original finding, the idea that what is damaged in visual neglect is one side of a display seems too simplistic²⁰; it does not account for the dissociation between visual and imaginal neglect (Beschin et al., 2000; Coslett, 1997), for the amodal nature of neglect (the deficit shows up in audition as well as vision, Marshall, 2001; Pavani, Ladavas, & Driver, 2002), for the fact that “neglected” stimuli typically provide some implicit information (Driver & Vuilleumier, 2001; McGlinchey-Berroth, Milberg, Verfaellie, & Grande, 1996; Schweinberger & Stief, 2001), for the characteristic response bias factors in neglect (Bisiach, Ricci, Luialdi, & Colombo, 1998; Vuilleumier & Rafal, 1999) and for the fact that higher-level strategic factors appear to play a central role in the neglect syndrome (Behrmann & Tipper, 1999b; Bisiach et al., 1998; Landis, 2000; Rode, Rossetti, & Bissoson, 2001). The “damaged display” view also does not account for the large number of cases of object-centred neglect (Behrmann & Tipper, 1999a; Tipper & Behrmann, 1996) in which patients neglect one “side” of an object viewed from an object-centered frame of reference (so that, for example, the ending of words are neglected regardless of whether they are on the left or the right or up or down). Moreover, as (Bartolomeo & Chokron, 2002) have documented (and reiterate in their commentary), the primary deficit in neglect is best viewed as the failure of stimuli on the neglect side to attract attention.

Unlike the tunnel imagery case described above, most cases of imaginal neglect are unlikely to be due to tacit knowledge. Deficits such as neglect, whether in vision or in imagery, represent a failure to orient to one side or the other. But a critical question is “one side or the other of what?” As in the example of S-R compatibility or imagery-induced perceptual motor adaptation, discussed in section 7.3.2, the direction of orientation may be better viewed as a direction in relation to objects in the world, rather than a direction in relation to objects in an image. Orienting is a world-directed response. When apparently attending to the left side of an image, patients may actually be orienting towards the left side of the perceived world (or perhaps of their body). Even with eyes closed we have accurate recall, at least for a short time, of the location of things in the world immediately around us and it may be that attention orients towards these world-locations. As I suggested earlier (section 7.3.2), it may be generally the case that it is the perception of physical space outside the head that gives imagery its apparent spatial character and that it does so by virtue of how mental contents are associated with (or bound to) objects or locations in the perceived world. The ability to bind objects of thought to the location of perceived (or recalled) external objects might allow us to orient to the objects in our image (using the visual index mechanism discussed in chapter 5, or the generalization of this mechanism I

²⁰ (Kosslyn, 1994) does not explicitly claim that the “depictive display” is damaged in cases of neglect, preferring instead to speak of the parallels between the vision and imagery systems. But to be consistent he *should* claim that the display is damaged, since one of the reasons for assuming a display is that it allows one to explain spatial properties of imagery by appealing to spatial properties of the display. Simply saying that it shows that vision and imagery use the same mechanisms does not confer any advantage to the depictive theory, since that claim can be made with equal naturalness by *any* theory of imagery format.

called Anchors, see note 17 and Pylyshyn, 1989). If something like this were true, it would provide a natural account of such parallels between image space and visual space as are raised by some of the neglect cases.

Despite studies such as those described above, the preponderance of clinical findings concerning mental imagery and vision show that the capacity for visual imagery and the capacity for vision are very often dissociated in cases of brain damage (see the review in Bartolomeo, 2002). This has been shown by the presence of normal imagery in patients with such visual deficits as cortical blindness (Chatterjee & Southwood, 1995; Dalman, Verhagen, & Huygen, 1997; Goldenberg et al., 1995; Shuren, Brott, Schefft, & 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 (Beschlin, Cocchini, Della Sala, & Logie, 1997, reviews the evidence for a double dissociation of neglect in vision and imagery). The case for independence of imagery and vision is made all the stronger by the extensive evidence that blind people show virtually all the skills and psychophysical phenomena associated with experiments on mental imagery, and may even report rather similar experiences of “seeing” shapes and textures as do sighted people. There is even some evidence suggesting that what characterizes patients who show a deficit on certain kinds of 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). On the other hand, insofar as blind people know (in a factual way) what objects are like (including aspects that are essential to their “appearance” – such as their shape, size, orientation, as well as other features that show up clearly in vision, such as smoothness) it is not surprising that they should exhibit some of the same psychophysical behaviors in relation to these properties. The pervasive evidence of the dissociation between imagery ability and visual ability is one of the most damaging sources of evidence for those who think that having a mental image involves using the visual system to “see” some state of the mind/brain.

7.5 What would it mean if all the neurophysiological claims turned out to be true?

Despite their problems, results such as those of Kosslyn et al. and Farah et al. (discussed above) have been widely interpreted as showing that topographical 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, 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. Since this idea comports with the experience we have that when we are imagining we are examining a display in our head, it has become the accepted view in cognitive science and even among philosophers who favor an empiricist view of mental states (e.g., Prinz, 2002). I have already suggested some reasons why the neuroscience evidence does not warrant such a strong conclusion (and that a weaker “functional space” conclusion is inadequate to support the claims of a special depictive form of representation for mental images).

If we are to take seriously the view proposed by picture theorists who take the literal cortical display interpretation of mental imagery, we need to understand the role that could possibly be played by such a literal picture on the surface of the visual cortex. Suppose that it is one day discovered that when people entertain a visual image there really is a picture displayed on the visual cortex and the picture has all the spatial and

depictive properties claimed for it (such as in the Kosslyn quote, reproduced in section 6.4.1). What would be the implications of such a finding? Would the claim that such a cortical picture is causally responsible for the phenomenology and psychophysics of mental imagery be consistent with the large body of experimental evidence regarding the function of mental imagery (e.g. all the evidence collected in Kosslyn, 1980; Kosslyn, 1983, 1994)? Equally important, what would we need to assume about the cognitive system (i.e., about the function carried out by the “mind’s eye”) for this sort of theory to work? What would the existence of a cortical display tell us about the nature and role of mental images in cognition? We have known at least since Descartes that there is a literal image on our retinas when we perceive, and early neuroanatomy suggested there is probably some transformed version of this very image on our visual cortex, yet knowing this did not make us any wiser about how vision works. Indeed, ruminating on the existence of such an image just raised problems such as why we do not see the world as upside down, given that the image on the retinas is upside down. It also led many psychologists to try to answer such questions as why we perceive objects to be roughly veridical in size despite differences in their retinal size, why we see a large colored panoramic view of the world when our eyes provide only a small peephole of high resolution colored information, why the world appears to be stable despite the rapid movement of our retinal image, and so on. These and other similar puzzles arising from the discrepancy between properties of our retinal image and properties of our percept led many people to assume that these discrepancies were overcome by presenting the mind with a corrected panoramic display, built up during visual perception. As we saw in Chapter 1, there is now overwhelming evidence against that proposal. This proposed panoramic display was a blind alley into which we were led by a strong tendency to reify our subjective impressions. The temptation to concretize a literal image in both vision and imagery, as well as the concomitant assumption of an equally literal “mind’s eye” may be very strong, but it leads us at every turn into blind alleys.

Even if it were shown that there is a picture displayed on the surface of the visual cortex during mental imagery, it would be hard to justify the leap from such a finding to a picture theory of mental imagery – one which explains the phenomenology and the psychophysics of mental imagery by the claim that people project a picture on their cortex and that they reason by visually examining this picture. I have already considered many reasons, both conceptual and empirical, to doubt the picture theory of mental imagery. Below I summarize a few of the reasons why a cortical-display version of the picture-theory is equally unsatisfactory.

7.5.1 The ‘mind’s eye’ must be very different from a real eye

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 the picture-theorists that cite this evidence.²¹ It not only suggests that the visual system is involved in imagery, and that it does so by examining 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 also similar to that of our eyes; it drops off with eccentricity and inscribes an elliptical acuity profile similar to that of our eyes (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, 1999). Since in the case of the eye, such properties arise from the structure of our retinas, it would appear to suggest that the mind’s eye is similarly constructed. Does the

²¹ In replying to this point, (Kosslyn, Thomson, & Ganis, 2002) says, “...these characteristics have been empirically demonstrated – which hardly seems a reason for embarrassment. Like it or not, that’s the way the studies came out. The depictive theories made such predictions, which were successful ...” Perhaps some people do not embarrass easily, but the fact that the data came out that way *ought* to be embarrassing to someone who believes that the data tell us about the format of images (or the architecture of the imagery system) since for these people it would entail that the “mind’s eye” is a duplicate of our real eye. It also ought to be embarrassing because the tacit knowledge explanation, which avoids this extravagant conclusion, provides a better account.

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 they did? Of course, the observed parallels could be just a 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 – i.e. to place themselves in certain circumstances and to consider what it would look like to see things, say, things located off 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 offered, 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 many of the properties of the peripheral parts of the visual system, are internalized in the imagery system. But if such properties were built in to the architecture, our mental 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 (if such a notion is even coherent) we would not be able to jump from place to place in extracting information from our mental image the way we can. And if images really were pictures on the cortex, the theory might well not have discharged the need for an intelligent agent to interpret them, notwithstanding claims that the system had been implemented on a computer (which is far from being the case, since a working model would require that all of vision be simulated). Even if there were a computer implementation of a high-performance system that behaved like a person who was examining a mental image, it would still not guarantee that what was claimed about the system, viewed as a model of the mind/brain, was true. As (Slezak, 1995) has pointed out, labels on boxes in a software flowchart (such as “visual buffer,” “attention window” or “pattern recognition”) constitute empirical claims that must be independently justified. They also constitute explanatory debts that have yet to be discharged, to use Dan Dennett's illuminating terminology (Dennett, 1978). Certain kinds of claims entail missing and unexplained intelligent processes, and the hyper-realistic “mind's eye” we are being invited to accept may well be such a claim.²²

7.5.2 The capacity for imagery is independent of the capacity for vision

The notion that viewing a mental image is very like seeing a scene must contend with a great deal of evidence showing that the capacity for visual imagery is independent of the capacity for visual perception (as we saw in section 7.4.3), and indeed, there is evidence for double-dissociations in the kind of damage observed in vision and in mental imagery (see the reviews in Bartolomeo, 2002; Beschin et al., 1997). If the early visual areas are the site of mental images and it is their topographical form that is responsible for the mental imagery results discussed earlier, it is hard to see why congenitally blind people produce the same imagery results (such as scanning times) as sighted people (Carpenter & Eisenberg, 1978; Zimler & Keenan, 1983). Cortically blind patients can also report imagery while some people who fail to report imagery have normal sight (Chatterjee & Southwood, 1995; Dalman et al., 1997; Goldenberg et al., 1995). Similarly, cerebral achromatopsia can be dissociated from the capacity to have colorful images (Bartolomeo et al., 1997; Shuren et al., 1996), hemispatial neglect can be manifested independently in vision and imagery (Beschín et

²² This example is reminiscent of a story I first heard from Hilary Putnam. It seems there was an inventor who bored his friends with unremitting stories about how he had invented a perpetual motion machine. When his friends stopped paying attention to him he insisted that he had actually constructed such a machine and would be happy to show it to them. So they followed him into his workshop where they were greeted by a wondrous array of mechanical devices, gears, levers, hydraulic pumps, and electrical bits and pieces. They stood looking at all this display in rapt amazement until one of them remarked to the inventor that, amazing though all this was, he noticed that nothing was actually moving. “Oh that”, replied the unruffled inventor, “I am awaiting delivery of one back-ordered little piece that fits right here and goes back and forth forever.”

al., 1997; Coslett, 1997; Guariglia, Padovani, Pantano, & Pizzamiglio, 1993) and visual agnosia can occur with intact mental imagery ability (Behrmann et al., 1994; Behrmann et al., 1992; Servos & Goodale, 1995). 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 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.

7.5.3 Images are not two-dimensional displays

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 (since the topographical mappings found in visual cortex are at best two-dimensional mappings of the retina). But even if a two-dimensional display was activated during episodes of imagining, it could not possibly correspond to what we mean by a mental image. The psychophysical evidence shows that mental images are, if anything, 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” – one of the most popular demonstrations of visual imagery – 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 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, as one would have to if the cortical display were to explain image-scanning and image-rotation results. Furthermore, images represent other properties besides spatial relations. For example, images must represent the motion of objects. People have assumed that once you have a way to “depict” relative locations you automatically have a way to depict motion. The naïve assumption is that motion is depicted as changing location, so that the motion of objects is represented by images that are themselves moving on the cortical display (although it is not clear what to do about motion in depth). But there is evidence that motion is encoded as a property distinct from location and change in location. For example people who are motion-blind or suffer from cerebral akinetopsia, are still able to tell that objects are in different locations at different times and can even detect motion in tactile and auditory modalities (see, for example, Heywood & Zihl, 1999).

Depicting translation and rotation requires depicting that the shape of rigid objects remains invariant. Consider the well-known example of mental rotation. What rotates in mental rotation? If the actual shape is to be depicted at a smooth succession of orientations, what ensures that the shape is retained – that parts move in a way that ensures that their location relative to other parts of the object remains fixed? In real rotation this is guaranteed by the object’s rigidity. Since nothing rigid is rotating in mental rotation, what enforces this rigidity constraint? What is it about a cortical display or a spatial (or depictive) format that makes rotation seem like a natural transformation, but other transformations seem odd (e.g., the transformation of a 3D object into its 3D mirror image, or its enantiomorph)? These questions have a natural answer in the represented domain – in the world of rigid objects. So perhaps what enforces the constraint is knowing that it holds in the world of rigid objects. Or maybe it is something about the architecture of early vision. What ought to be clear is that the maintenance of the shape of a figure as it rotates (as well as the figure’s passage through intermediate locations) does not follow from the mere fact of it being represented on

a 2D surface. Yet people continue to cite the spatial nature of image representations to explain mental rotation.²³

Mental images represent not only spatial properties. They also represent the color and luminance and shape of objects. Are these also to be found displayed literally 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?

7.5.4 Images are not retinotopic

The cortical display view of mental imagery assumes that mental images consist in the activation of a pattern that is the same as the pattern activated by the corresponding visual percept. It follows then that such a pattern corresponds to the retinotopic projection of a corresponding visual scene. This assumption is a direct consequence of viewing a mental image as the same as the visual image that is believed to occur in visual perception, which is why the finding of (Tootell et al., 1982) is always cited in discussions of the nature of mental images (see section 7.4.2, especially Figure 7-2). But such a retinotopic pattern does not correspond to a mental image as the latter is understood in the psychological literature, and it is not what we “see” in our “mind’s eye”. A mental image covers a much larger region than the fovea, and may even cover the region behind the head (as Attneave suggests in the quotation reproduced in section 7.3.2). David Ingle has also pointed out (Ingle, 2002) that since mental images (particularly memory images) typically remain fixed in allocentric coordinates, they must correspond to processes further along the visual pathway than areas 17 or 18. Cells whose receptive fields remain fixed in allocentric coordinates are found in inferotemporal cortex or in parietal cortex, but these areas are not mapped in terms of a 2D map the way cells are in visual cortex. Moreover, Ingle argues, since mental images contain recognized and localized objects, images must be located after the “two visual systems” converge, such as in prefrontal cortex, where there is no evidence for a topographical organization.

There is some inconsistency in how picture theorists describe the cortical display. On the one hand the only evidence for a clearly topographical representational structure in cortex is the retinotopic structure in early vision. Information higher up in the visual stream tends not to be topographically organized (at least in a way that reflects visual space). Consequently, proponents of the cortical display view cite evidence such as presented by Tootell et al., as well as evidence of activity in the retinotopically mapped visual cortex during mental imagery. On the other hand, among the reasons put forward for the existence of a cortical display (Kosslyn, 1994, Chapter 4), is that it is needed to explain the stability of our perception during eye movements and the invariance of recognition with movements of the retinal image (Kosslyn & Sussman, 1995, also assumes that amodal completion and “filling in” occur in this cortical display). But for that to be the case, the image would have to be panoramic rather than retinotopic – it would have to display the larger stable view constructed from the sequence of saccades. Unfortunately there is no evidence at all to support that kind of cortical display. Indeed, as we saw in section 1.4, there is every reason to believe that vision does not achieve stability and completeness by accumulating information in a spatially extended inner display. There is a great

²³ For example, in his critical discussion of the idea that representations underlying mental images may be “propositional” (Prinz, 2002, p 118) says, “If visual-image rotation uses a spatial medium of the kind that Kosslyn envisions, then images must traverse intermediate positions when they rotate from one position to another. The propositional system can be designed to represent intermediate positions during rotation, but that is not obligatory. If we assume that a spatial medium is used for imagery, we can predict the response latency for mental rotation tasks...” But Prinz does not tell us why an it is obligatory in a “spatial medium” that “images must traverse intermediate positions” or what he thinks is passing through intermediate positions in the mental rotation task and why it is obligatory that it do so. Passing through intermediate position is obligatory if an actual rigid object is moving or rotating. If a representation of an object, or an experience (a phenomenal object) is rotating, the laws of physics do not apply so nothing is obligatory except by stipulation. And stipulating that the mental image must pass through intermediate positions occurs only because its referent (a real rigid object) would have done so. It seems that even philosophers are caught up in the intentional fallacy!

deal of evidence showing that visual stability and saccadic integration is not mediated by any kind of inner-display (Blackmore, Brelstaff, Nelson, & Troscianko, 1995; Irwin, 1991; McConkie & Currie, 1996; O'Regan, 1992). For example, information from successive fixations cannot be superimposed in a central image as required by this view. Recent evidence also shows that there is no central repository where visual information is enriched and accumulated to form a detailed panoramic view of a scene, of the sort we typically experience. For example work on change blindness shows that the visual system stores very little information about a scene between fixations, unless attention has been drawn to it (Rensink, 2000; Rensink, O'Regan, & Clark, 1997; Simons, 1996). Thus it appears that there is no panoramic pictorial display. On the other hand, the proposal that mental images are displayed on a retinotopic display, while consistent with neuroanatomical data, is inconsistent with the findings of mental imagery experiments as well as with the phenomenology of mental imagery. Both these sources of data conflict with the assumption that only retinotopic (and particularly foveal) information is displayed. Thus there appears to be little support for either a retinotopic or a panoramic display as the basis for mental imagery.

7.5.5 Images do not provide inputs to the visuomotor system

I have suggested that mental images get their spatial character because the objects that one is imagining are bound (by means of indexes) to perceived locations in real space, including information from proprioception and other perceptual modalities. As a result, imagined locations are tied to real space and hence provide the basis for a certain kind of spatio-motor coordination. According to the account I have been giving, the only thing that engages the motor system are the perceived locations to which imagined objects are bound (and that's what the location of objects in an image *means*). A more interesting question is whether anything more than location is involved when images interact with the motor system. For example, is there is any evidence to suggest that an image can engage the motor system in terms of more detailed properties of the image content, the way that real vision engages the motor system?

When we look in detail at cases that involve more than just the location of imagined objects, we find that images do not interact with the perceptual-motor system in a way that is characteristic of visual interaction with the perceptual motor system. To show this we need to examine certain signature properties of the visual control of movements, rather than at cases where the control may actually be mediated by spatial attention or visual indexing of the sort introduced in chapter 5. One clear example of the strictly visual control of motor action is the type of eye movement called *smooth pursuit*. People can track the motion of slowly moving objects with a characteristic smooth movement of the eyes known as *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 by smooth pursuit (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 surrounding frame (Wyatt & Pola, 1979). In other words they can engage in the smooth pursuit of inputs generated by the early vision system and perhaps the proprioceptive system as well. 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).

There are also significant differences between the way that the other parts of the motor system interact with vision and the way they interact with mental images. Consider the visual control of reaching and grasping. Although we can reach out to grasp an imagined object, when we do so we are essentially reaching towards a location and mimicking a grasping gesture. The movement we execute resembles a *pantomiming*

movement rather than a movement generated under visual control. The latter 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. Such differences provide strong reasons to doubt that imagery can serve as input into the dorsal stream of the early vision system, where the visuomotor control process begins. What the evidence does support, however, is the notion that the *locations* of imagined objects can have observable consequences on behavior. But these results are best accommodated by the independently motivated assumption that observers can associate (bind) imagined objects to real perceived objects or orientations using either visual indexes (FINSTs), when they can see the environment, or proprioceptive indexes (*Anchors*) when their eyes are closed. This proposal was discussed at length in sections 7.3 and 7.3.2.

There is considerable evidence that the visuomotor system is itself an encapsulated system (Milner & Goodale, 1995) which, like the early visual system, is able to respond only to information arriving from the eyes, often including 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 this system the way that visual inputs do. In the next section I suggest other reasons to doubt that the early visual system interprets visual images as though they were actual pictorial displays.

7.5.6 Examining a mental image is very different from perceiving a display

In accessing information from a real scene, we have the freedom to examine it in any order, and may even access some of it in parallel. But this is not true of accessing information from mental images. Take the following simple examples. Imagine a familiar printed word (e.g., your name) and try reading or spelling it backwards from your image. Write down a 3 x 3 matrix of random letters and read them in various orders. Now memorize the matrix and try doing the same from your image of the matrix. Unlike in vision, some orders (e.g., the diagonal of the matrix read 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 a mental image generated from memory, such as assuming a limit on the number of elements that can be drawn, or assuming that elements decay. Such assumptions are completely ad hoc. For example, in many imagery studies visual information is not found to fade rapidly (Ishai & Sagi, 1995) nor does it appear to fade in the case of images used to investigate mental scanning phenomena (which, like the map used by Kosslyn et al., 1978, is more complex than a 3 x 3 matrix of letters). 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 structure of the image that determines how the image can be read and manipulated (i.e., to account for the fact that it is how one interprets the image, rather than its geometry, that determines its apparent fading). For example, it is how figures are conceptualized that determines the difficulty of an image superposition task (illustrated in Figure 1-19f), or to how quickly figures can be "mentally rotated" (illustrated in Figure 6-8).

The fact that mental images represent the conceptual content of a scene (either recalled from memory or constructed during certain tasks) explains why images are distorted or transformed over time in characteristic

ways (see, the examples in section 1.4.3), why mental images can't be visually (re)interpreted, and why they can fail to be determinate in ways that no picture can fail to be determinate (Pylyshyn, 1973, 1978). 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. Imagine throwing a ball in the air; then ask yourself about incidental perceptual properties of the event, such as the color or weight of the ball and whether it was spinning, the appearance of the background against which you saw the ball rise, how long it took to reach the peak of its trajectory or to fall back to earth, and so on. When recollected images are incorrect, they are invariably incorrect in conceptual ways (some discrete object is missing or mislocated or its properties are incorrectly conjoined); an incorrectly recalled image is not like a photograph that has an arbitrary piece torn off.

Not surprisingly, there are also many ways of patching up a picture theory to accommodate the fact that image contents are interpreted rather than strictly iconic. For example one can allow images to be tagged as having certain properties (perhaps including the property of not being based on real perception), or one can assume that parts of images have to be refreshed from time to time from conceptual representations stored in memory, thus bringing in conceptual factors through an image generation function. With each of these accommodations, however, the depictive format has less and less of an explanatory role to play because the work is being done elsewhere. It becomes like an animated computer display whose behavior is determined by an extrinsic encoding of the principles that govern the animation, rather than by intrinsic properties of the display itself.

There are two basic problems with the assumption that the contents of a mental image are like the contents of a picture, and therefore not conceptualized. The *first* problem is that both the contents of an image and the dynamic properties that an image has can be whatever you wish them to be, and the fact that you have a choice means that you must at some stage have had some conceptualized content in mind – you must have had an interpretation in mind. Thus no matter what content you put into your image representation, you do so under a particular description or conceptualization or intended interpretation. For example, if you decide to imagine a rectangle as such, you know that whatever sort of representation you construct you do so with the knowledge that it is meant to be a “rectangle” so you couldn't mistake it for, say, a square or a parallelogram – as you might if it were a real uninterpreted form. And you couldn't mistake your image of a wire cube for its perceptually reversed interpretation, or your image of two parallelograms above one another with corresponding vertices joined as a Necker cube (as in the example discussed in section 6.5.4), even though in each case you might if you drew it.

There is, however, one caveat to the above claims about the content of images. The choice of image content is subject to the understanding that to “mentally image” something generally means to imagine that you are seeing it. Consequently, this means that certain aspects will tend to be assigned some value or other (as opposed to being left indefinite). For example, when asked to “imagine a printed word” it would be reasonable (though not obligatory) for you to imagine a sequence of letter-shapes, which in turn invites you to choose upper or lower case letters and maybe even a particular font. None of these choices is mandatory (and indefinitely many visual properties will not get assigned a value – e.g., the nature of the background on which the letters are imagined), yet failing to assign some shape properties may tend to suggest that you are not following the instruction (including the self-instruction) to *imagine seeing* a printed word (after all, if you saw a printed word it would have letters with particular shapes).

The *second* and related problem is that no display is informationally-equivalent to a *mental image*. This is

a point I discussed earlier (in section 6.5.5) in connection with Shepard's conjecture that the mental state corresponding to having a mental image could in principal be externalized as a picture, in such a way that a person who saw the picture would be in a similar mental state as the person who had the mental image. The reason this would not work is that no visual stimulus carries information about the picture's interpretation. By contrast, mental images *are* the interpretation and are therefore conceptual. As (Kosslyn, 1994, p 329) put it, mental images contain "previously digested information." What is even more significant is that there is no reason to believe that they contain anything else.

The above examples illustrate a very general problem with the view that examining an image is just like seeing a display. A few other examples where this parallel fails are outlined below.

1. *Images do not have the signature properties of early vision.* Because vision (at least early vision) is encapsulated, it works according to different principles from those of rational reasoning. It provides interpretations of visual patterns according to certain rules, such as those I listed in section 3.1.1.2. In particular it has signature properties such as the properties discussed in (Hochberg, 1968) – e.g., certain line drawings are automatically interpreted as three dimensional constructions, certain patterns are ambiguous and result in involuntary reversals or changes in perspective, and certain sequences of images lead to the automatic perception of motion. If we create a mental image from a description (as in the example I presented in section 6.5.4 where I described a Necker Cube, in terms of two identical joined parallelograms) we do not find such phenomena as spontaneous interpretation of these 2D shapes as 3D objects, spontaneous reversals of bistable figures, amodal completion or subjective contours (Slezak, 1995), visual illusions, or the incremental construction of visual interpretations and reinterpretations over time, as different aspects are noticed. This is just what one would expect if the mental image was an interpretation of some possible scene, as opposed to an uninterpreted image of one.
2. *Perception of an actual retinotopic pattern is different from having a mental image.* An interesting test would be if one could compare the case where a pattern was projected onto the cortical display with the case where one merely imagined the same pattern. This is not an experiment that we are in a position to carry out at this time. But we can come close. Because of the retinotopic mapping from retina to primary visual cortex we should be able to create a known pattern of activity on the cortex by projecting the pattern on the retina. We can do that, without running into problems with eye movements, by creating an afterimage on the retina. When we do that we find an interesting difference between the visual appearance of information projected onto the retinotopic cortical display and the appearance of the same pattern when it occurs 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. The farther away the background is, the larger does an afterimage appear. By contrast, mental images imagined over a perceived scene do not change their apparent size depending on the distance of the background, providing strong evidence that mental images are not identical to images projected onto the retinotopic layers of the cortex. (Compare the Emmert's Law prediction based on the assumption that images are displayed on the visual cortex, with the claim I presented earlier, concerning what happens when we "project" a mental image onto a real

scene. I claimed that what we do not superimpose two images, but rather we simply think of the objects in our image as being located at places in the world occupied by some visible feature or object that we pick out and index. If the indexing view is correct, then we would not expect the distance of the background to make any difference initially, since we are free to pick features whose locations correspond to where we want certain imagined elements to be. But once we have locked our imagined elements onto features in this way, if the background begins to recede then we might expect the apparent size of the image to shrink as the distance between indexed features diminishes. So I would not expect a mental image to obey Emmert's law as the background recedes. If anything, I might expect the opposite: Emmert's law says that a retinal/cortical image looks bigger on a more distance background whereas the indexing view says that the size of the image shrinks because the distance between indexed elements diminishes as the background recedes. Of course I also maintain that your image has whatever properties you wish it to have, so perhaps your belief in the fixed size of the imagined scene overrides the shrinking distances between indexed elements. In either case I clearly do not expect Emmert's law to hold for mental images.)

It appears that if having a mental image consists in certain activity in the visual cortex, this activity does not function as a *display* of that image, in the sense in which a picture or a TV screen is such a display. In particular, a pattern of cortical activity is not something that must be visually interpreted, the way that the electronic activity on the TV screen must be visually interpreted. As to whether the activity is in some other way similar to the activity involved in visual perception, this remains an open question. The assumption that is being questioned is not whether entertaining mental images have *some* connection with visual perception, it is the assumption that entertaining a mental image consists in visually interpreting some special form of representation, in particular a form of representation that is picture-like or, as some people prefer to call it "depictive." Although the term depictive is not well defined beyond its metaphorical connection to pictures, it is intended to suggest that the image is pre-perceptual or pre-conceptual, and therefore in need of interpretation by the usual early visual processes. This is why people have been concerned to show that activity associated with mental imagery occurs in areas of the cortex known to be retinotopically mapped. But, as I argued earlier, just because there is activity in an area that is retinotopically mapped is not sufficient reason conclude either that the mental images themselves are retinotopically mapped in a way the preserves spatial properties, or that if they are, that they serve as input to a visual system which perceives them in a way determined by their retinotopic shape. It is these assumptions, rather than the mere involvement of parts of the visual system, have been the contentious assumptions in the mental imagery debate, at least since (Pylyshyn, 1973).

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

Where, then, does the “imagery debate” stand at present? That all depends on what you think the debate is about. If it is supposed to be about whether reasoning using mental imagery is somehow different from reasoning without it, who can doubt that it is? If it is about whether in some sense imagery involves the visual system, the answer there too must be *yes*, 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 big open question is; *in what way* is the visual system involved? Answering that is likely to require a better functional taxonomy of the visual system and better alternative proposals for how non-deductive (image-based) reasoning might proceed. 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 topographically-faithful remapping 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 accurate representation of geometrical, spatial, metrical or visual properties depictive? If that makes it depictive then any precise description of how something looks is thereby depictive. Does being depictive require that the form of the representation be spatial? As I have suggested, that depends on what restrictions are placed on “being spatial.” Does being spatial 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 faithfully *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 (especially an extensible one such as the floating point real number system, or the Dewey Decimal system used in libraries), as well as any analogue medium, can represent magnitudes as precisely as needed. If the claim that images preserve metrical spatial information means that an image *uses spatial magnitudes to represent spatial magnitudes* (by mapping magnitudes in a monotonic manner), then this is a form of the literal picture theory that I have argued is incompatible with the available evidence.

The neuropsychological evidence I have briefly examined, 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. One major problem is that we are attempting to account not only for certain behavioral and neurophysiological facts, 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 history of other sciences and even of other parts of cognitive science is any indication, the eventual theory is likely to be at odds with 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.

The typical response I have received to arguments such as those raised in this chapter is that the critique takes the picture theory too literally and nobody really believes that there are actual pictures in the brain. Almost every article I have seen by advocates of the picture theory is at pains to point out that images are not in every way like pictures. For example, (Kosslyn, 1994, p329) states, “images contain ‘previously digested’ information” and (Denis & Kosslyn, 1999) state that “No claim was made that visual images themselves have spatial extent, or that they occupy metrically defined portions of the brain.” But then how do they explain the

increase time to scan greater image distances or to report details in smaller images? The explanations of these phenomena *require a literal sense of 'spatial extent'* otherwise the depictive theory is indistinguishable from what I have called the null hypothesis (see the discussion of the 'functional space' alternative in section 7.1). And how, if one forswears the literal view of a cortical display, is one supposed to interpret the concern about whether imagery activates topographical areas of the visual cortex, or the claim that such activation establishes that images, unlike other forms of representation, are "depictive"? 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 story. This is the strategy of claiming a decisive advantage for the depictive theory because it has the properties referred to in the quotation in section 6.4.1, is located in the topographically organized areas of visual cortex, "preserves metrical information" and so on, and then, in the face of its implausibility, systematically retreating from the part of the claim that is doing the work – its literal spatial layout. As Bertrand Russell (Russell, 1918/1985, p 71) once said about the advantage of postulating what you would like to prove, such a strategy "...has many advantages; they are the same as the advantages of theft over honest toil."

7.6 What, if anything, is special about mental imagery?

Notwithstanding the skepticism displayed in this book concerning many contemporary views of vision and mental imagery, I believe that visual science has made considerable progress in the past few decades. The trap we have had trouble avoiding is that of taking our introspections at face value, as showing us what vision is like and what its products are. Whatever the representations generated by vision may be like, they may well be the same, or at least very similar to, those generated when we are engaged in visual imagining. And those representations may well be different from those generated when we reason about more abstract ideas. What we still do not have is a good idea of how such perceptually based representations are different from ones that are not accompanied by sensory experiences (though we have had numerous inadequate ideas). It may be worth speculating on ways in which these may be different. But first let me review and try to be as explicit as I can about what it is that I do (and do not) claim about mental images.

7.6.1 What I am *not* claiming: Some misconceptions about objections to the picture-theory

There has been a great deal of misunderstanding of the position that I have taken in the past on the nature of thought and of mental imagery (e.g., Pylyshyn, 1973, 1981, 1984b). For example, here are a few of the claims that people have attempted to foist on those of us who have been critical of the phenomenologically-inspired "picture theories" of vision, and more particularly, of mental imagery. The following are some beliefs about mental imagery that I do not hold.

1) *Mental images don't really exist; they are merely "epiphenomenal."* Prior to the development of an adequate theory of reasoning using mental imagery, the term "mental image" is only the name for the experience of "seeing" without the presence of the object being seen. Nobody can deny the existence of such an experience. Moreover, the content of the experience is a piece of data that we rely on, along with a great many other sources of evidence, in formulating a theory of the what goes on during certain kinds of episodes of information processing. Such episodes are typically, though not necessarily, accompanied by the experience of "seeing" in one's "mind's eye". In any case the claim that mental images are epiphenomenal is at the very least ambiguous: It corresponds either to the claim that we are deluded about what we experience

– that we do not experience an image as something that is “seen”, which is absurd, or to the claim that a scientific theory of mental imagery will not incorporate things that are like what we experience when we have a mental image, which is almost certainly true. Our experience is not the experience of seeing a mental event, but rather it is the experience of seeing a possible perceptual world. Consequently no theory of information processing will mirror our phenomenal experience by hypothesizing the existence of objects inside the brain that are the same as the objects of our experience. Nor will a neurological theory do so. That’s because the objects of our experience are things in the world (perhaps nonexistent things, but nonetheless things whose proper place is in a world outside the head). As Ned Block has correctly pointed out (Block, 1981a), the appeal to epiphenomenalism is either just another way of stating the disagreement about the nature of mental images, or it simply confuses the functional or theoretical construct “mental image” with the experience of having a mental image.

2) *The form of representations underlying mental images is propositional.* It is certainly possible (and even quite plausible) that the content of **both** images and visual percepts can be adequately encoded in a propositional²⁴ (or more precisely, quasi-sentential or quasi-logical, or symbolic) form of some sort. Indeed, such an encoding would help explain the conceptual properties exhibit not only by mental images, but also by perception itself (as I suggested, for example, in Pylyshyn, 1973, as well as in Chapter 1). But until someone produces a detailed theory that accounts for some significant imagery phenomena using a propositional encoding, this proposal serves primarily to illustrate some of constraints on an adequate form of representation. Thus an important use of the idea that images are encoded in propositional form is as a *null hypothesis* against which to test various proposals for image representations. There is something special about propositional representations. Thanks to the work in formal logic and computation theory over the last half-century we know some important things about what have been called formal languages that we don’t know about other forms of representation. Such propositional symbolic representations have a well-defined combinatorial syntax, together with a semantic theory and a proof theory and, if not a theory of other types of inference (e.g., inductive, abductive, heuristic), at least some indication of how semantic properties might be preserved over certain syntactic transformations. In other words, we know some of the essential formal (semantic and syntactic) properties of logical calculi, as ways of encoding propositions. Consequently such a system of encoding constitutes an appropriate null-hypothesis against which to compare other theories of the encoding of mental images. When a theory of mental imagery is proposed we can ask: Can it explain anything that a quasi-linguistic theory would not be able to explain (and vice versa)? If not, nothing is gained by accepting the proposal and a great deal is lost, such as the entire apparatus of inference as developed in formal logic and continued in non-classical logics such as the non-monotonic logics studied in artificial intelligence.

3) *There is nothing special about the representations underlying visual imagery – they are the same as representations for verbal or other forms of reasoning.* Putting aside the difficulty of distinguishing different “types of representation,” there appears to be every reason to think that *something* is special about the sort of reasoning that is accompanied by the experience of mental imagery – something that distinguishes it from other forms of reasoning. The relevant question to ask about some particular proposal for what is special about mental images is not whether it comports with our intuitions, but whether it can be empirically sustained. And here I do claim that every concrete proposal I have seen of what constitutes this special ingredient that is present in the case of mental images and absent from other forms of thought has been either demonstrably false or else it has been ambiguous or incoherent (or both). The proposals have typically been along such lines as that images are picture-like or that they “preserve the metrical properties of the objects they depict.” As I have argued (Pylyshyn, 1981), such claims have invariably been presented in what

²⁴ I am using the term “propositional” in its common, though somewhat informal sense to mean any language-like symbolic encoding for which a formal syntax and rules of inference are available and which has a truth value (i.e., propositions are either true or false, or something else if you have a multi-valued or fuzzy logic). In other words, by “proposition” I really mean a statement in a logical calculus of some sort, such as the predicate calculus.

amounts to an intellectual shell game, in which images are claimed to *have* metrical properties or to be pictures when the theory needs to explain certain metrical correlates of represented magnitudes (such as the longer time it takes to scan greater image distances), and merely to *represent* metrical properties when the literal picture view seems untenable. The difference between *having* metrical properties and *representing* them is fundamental: the first is a claim about the form or the physical property of the system of codes used to represent magnitudes such as distance, while the second is simply a claim concerning what properties of the world can be represented (i.e., it is a claim about the representational capacity of the system of codes). The second is obviously much weaker than the first since, for example, it is fulfilled universally by human languages (since all languages spoken in industrialized parts of the world have a productive system of names for magnitudes, such as the numerals). The first claim (that images have metrical properties) can be taken either as a formal mathematical claim (i.e., the system of codes has a formal property that supports the metrical axioms), which is often put in terms of the claim that images are encoded as analogues, or as a literal claim that images are laid out in physical space in the brain. Since an analogue theory of imagery (or of space) has yet to be developed I have concentrated entirely on the picture-theory alternative in the last two chapters. Also, in view of the cognitive penetrability of most imagery phenomena, there is good reason to doubt that an all-analogue representation of imagery will do (as we saw in section 7.1).

4) *The visual system is not involved in manipulating and interpreting mental images.* The claim that imagery and vision are closely linked and use the same mechanisms may or may not be true, or even meaningful, depending on how broadly one takes the notion of “visual system.” If by “visual system” one means any process concerned with interpreting visual information, including the processes involved in allocating attention and recognizing visual objects, then the claim is almost certainly true. But it is true for the uninteresting reason that much of what happens in this extended sense of vision is a form of reasoning so it will naturally also apply to reasoning using mental images. If, however, the claim is that what I have called “early vision” – the part of the visual system, discussed in Chapter 2, that is unique to vision and is cognitively impenetrable – then there is good reason to doubt that these mechanisms are involved in the examination of mental images. I examined some of the evidence for this claim in Chapter 6.

5) *The so-called “imagery debate” is over and one or the other side is the winner.* The trouble with this statement is that although there has been a great deal of discussion about the nature of the representations and mechanisms involved in mental imagery, such discussions hardly qualify as a debate since there are not two (or more) well-defined sides, due largely to the fact that most theories of mental imagery rest on undefined (and perhaps undefinable) notions. The “debate,” so far as I understand it, is not between two theories or even two classes of theories of mental imagery. Rather it has been about whether *particular* proposals that have been put forward have any empirical validity (or in some cases, conceptual coherence).

The focus of the discussion about mental imagery has changed in the past decade and there has been a great deal more evidence brought into the discussion, especially evidence from clinical neurology and from various forms of brain imaging (such as PET, MRI, fMRI) and recently also from an intrusive technique of disabling part of the brain using transcranial magnetic stimulation (rTMS). Yet this increase in the empirical base, welcome though it is, cannot offset the fact that the questions and options under discussion continue to be ill-formed. No matter how much objective reliable evidence is introduced, the “debate” cannot be resolved until the nature of the claims is made clear. There is, at the present time, an enormous gap between the evidence, the models, and the claims being made about the real nature of mental images.

7.6.2 What constraints should be met by a theory of mental imagery?

In the absence of a worked-out theory of mental imagery it is useful to consider some plausible constraints on representations that might underlie images. The following speculative list includes some of the ideas I have discussed throughout this and the previous chapter:

1. *Image representations contain information about the appearance of things*, so they use the vocabulary of visual properties (e.g., color, brightness, shape, texture, and so on). This is a very different claim from the empiricist claim, going back to Hume, that the vocabulary *consists of sensations*; a claim recently revived and repackaged in modern dress by many writers (Barsalou, 1999; Prinz, 2002).
2. *Image representations contain information about the relative location of things*, so they use the vocabulary of geometrical relations (above, inside, beside, to-the-right-of, and so on). Indeed, there is reason to think that representing spatial properties is a more fundamental characteristic of what we call images than is representing appearance properties (Farah, Hammond, Levine, & Calvanio, 1988), but that this can be done without invoking a spatial medium, by relying on the spatial properties of immediately perceived space.
3. *Image representations typically refer to individual things*; they represent *token* individuals (things or objects, or whatever early vision delivers). The content of an image may also ascribe properties to these tokens, but they generally do not contain quantified assertions. For example they might assert $Q(x_1)$, $Q(x_2)$, ... $Q(x_n)$ for n distinct individual x 's instead of the general assertion "There are n things that have property Q ." Image representations can, however, represent abstract properties, such as that X *caused* Y , something that no picture could do in a direct manner (without some iconic convention which would, in effect, serve to label parts of the image). Since a mental image constitutes an interpretation it can represent abstract properties without such conventions simply because the person who has the image knows the intention or the conceptualization under which it was created. (This however means that part of what an image denotes is offstage in the intention and does not appear in the part of the representation experienced as a picture. This fact plays a role in the claim that I will discuss in section 8.2.3, that one cannot think in images alone.)
4. *Image representations lack explicit quantifiers*. Images are not able to express such assertions as that *all* things in an image are X 's, or that *some* of the things in the scene are Y 's and so on, although they can represent some or all individuals in the image as being X 's. Representing a content that is quantified (there is an unspecified individual who has property P , or all individuals that have property P also have property Q) can only be accomplished by adding symbolic tags that may be interpreted to have the meaning of a quantifier – which is to say, an image *qua image* can only represent quantified or generalized properties by means that are essentially non-pictorial.
5. *Image representations lack explicit disjunctions*. Images are not able to express such assertions as that either individual X or individual Y has property P , or that individual Y has either property P or property Q . The closest one can come to this is to have two images, one with only X and one with only Y .
6. *Image representations lack explicit negation*. Images cannot carry the information that there are no X 's in the image or that none of the X 's have property P . Rather, negation is expressed indirectly by the absence of certain things in the image, together with the implicit assumption that all relevant things are included.

7. *Image representations provide access paths* that make getting from one item or property to another more direct for some items and properties than for others. Relations such as *adjacent* may have a privileged position in terms of providing more direct access paths than, say, *same-size-as*, so that it is easier to get a locus of processing from one individual to an individual that is represented as adjacent to it, than to another individual that is the same size.
8. *Image representations may lend themselves to certain kinds of transformations* in preference to other kinds of transformations. Without assuming any particular format for images, it might be that carrying out certain transformations on them is more natural than other transformation. This may have to do with which properties are encoded (e.g., orientation may be encoded independently of shape), or it may have to do with computational complexity issues. It may also be simpler to carry out certain operations in a certain sequence. For example, in order to compare two shapes that differ in their orientations, it might be computationally cheaper (or even necessary due to limited resources) to go through the computation of the representation of the shape at intermediate orientations (see section 6.3.2.4 for a brief discussion of why this might be so).
9. *The information about individuals in a mental image can be associated with individuals in a perceived scene.* This is what allows images to inherit spatial properties of visually perceived scenes. As I suggested in section 7.3, we can think about particular imagined individuals and use indexes to bind them to objects in a scene we are viewing. Doing so would keep their relative locations fixed so long as the background scene was fixed and rigid and would ensure that other implicit relations among bound objects held (e.g., if three imagined objects were bound to three collinear scene objects, the location of the middle imagined object could be visually perceived to be “between” the other two imagined objects.) In section 7.3.2 I proposed that something like this might even be possible when the scene is not perceived visually, but perceived in some other modality (e.g. acoustically, proprioceptively or kinesthetically). This simple assumption allows us to explain the apparent ability to project an image onto a scene so that the combined image-and-scene behaves in certain ways like a superimposed image-percept scene.

With a little thought this list could be extended. These proposed characteristics or constraints on mental images are quite different from those proposed by picture-theorists. None of the theoretical ideas about the nature of mental imagery in cognitive science take these desiderata seriously. Mental pictures do have some of the properties on this list, but they also have serious drawbacks. The only theory I am aware of that deals with this aspect of reasoning from images, and that appears, *prima facie*, to have some properties listed above, is a system of formal representation proposed by (Levesque, 1986). Levesque describes an expressively weaker, but more efficient, form of logic that he refers to as a “Vivid Representation.” This proposal has the merit of recognizing that we can have forms of representation that are more limited in what they can express but have the special feature that they allow certain conclusions to be drawn rapidly – essentially by a form of pattern-matching. Like images, they do not allow one to directly express negation (e.g., the only way that they can represent the proposition “there are no red squares in the scene” 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, to both be treated as true), and they do not allow quantification (e.g. they can only represent the proposition “all squares are red” by explicitly representing each square, however many there are, and asserting of each one that it is red). Like images, they cannot express the fact that there are 5 objects in the scene, except by representing each of the objects, of which there would have to be 5 in all. I

had also informally proposed a similar set of ideas in speculating about what might be special about representations underlying imagery, as opposed to representations underlying other kinds of thoughts (Pylyshyn, 1978). These are, admittedly, small steps towards a formalism for representing mental images. They do not suggest why such representations should be accompanied by the experience of seeing, although they do have the virtue of being limited in some of the same ways that images are.

8. SEEING WITH THE MIND'S EYE 3: VISUAL THINKING

8.1 Different “styles” of thinking

One of the most widely accepted ideas about reasoning and problem-solving is that there are different styles of thinking and that these differing styles are most clearly characterized in terms of whether people are “visual” or “verbal” thinkers (or sometimes “concrete” or “abstract” thinkers). There is little doubt that people differ in a number of ways with respect to their habitual or preferred styles of thought and their approach to problem solving. There is surely also something to the observation that some people tend to think in a way that is in some sense more “visual” in that their thoughts more often concern the appearance of things and may be accompanied by vision-like experiences. The problem is in spelling out this difference in a way that takes it out of the realm of personal experience and connects it with a scientific theory. Describing some people’s style of thinking as “visual” may imply any one of several things; it may suggest that they prefer to use visual aids (models, mock-ups, diagrams, etc) when they solve problems; or that they prefer to solve problems that are related to how something looks, or that that they use their visual skill in some internal way (presumably by working with mental images) even when they don’t have something concrete to look at.

It is perfectly reasonable to say that some people prefer to think about appearances, just as it is perfectly reasonable to say that some people prefer to talk about appearances. But that tells you nothing about the format of their thoughts. Pictures presumably depict appearances. But sentences such as “he looks tired” or “she looks good in her new hairdo” or “the big red ball is behind the blue door” also refer to appearances. Terms like “red” refer to a visual property. But these are all properties of things being described, not of sentences or of thoughts or of mental images. It is the content that is visual, not the form itself. If I am right that thought does not in any way consist in examining mental pictures, then neither does it consist in listening to mental sentences in an inner dialogue, as introspection suggests. Here we come, once again, to the conflicting demands of one’s conscious experience and those of a scientific theory that has explanatory power. It is to this conflict that I now turn.

8.2 Form and content of thoughts: What we think with and what we think about

8.2.1 The illusion that we experience the *form* of our thoughts

Ever since the use of mentalistic terms (such as “know”, “believe”, “think”, “want” and so on) became once again permissible in psychology and philosophy, after a long dry period of behaviorism, people have found it completely natural to assume that we have conscious access not only to the content of our thoughts, but also to the form that they take. Thus we find it natural to suppose that our thoughts take the form of either inner dialogue (thinking in words) or of imagining a visual scene (seeing an image in our “mind’s eye”).

²⁵ In fact this homily has been translated into what is known as the “Dual Code” theory of mental representations (which was developed most extensively in the work of Paivio, 1986). The dual code idea

²⁵ Notice that when you experience something, you invariably experience it as something *sensory*. Try imagining an experience that is not the experience of perceiving something – of seeing, hearing, feeling, smelling or tasting. Even inner feelings like pain or queasiness, or emotions such as anger, are experienced in terms of some perceptual sensations (in fact the James-Lange theory of emotions equates emotions to perceived bodily states). Given this universal property of conscious experience (for which there is no explanation), it is no surprise that when we experience our thoughts we experience them as seeing or hearing or some other perceptual modality (touch, taste, proprioception): What else *could* we experience them as?

infects our interpretation of self-reports of thought. Most people – including great thinkers like Einstein, Maxwell, Faraday, Helmholtz, Galton, Watt, Tesla and others (see, for example, the review in Shepard, 1978a) – maintain that their deepest thoughts are “devoid of words.” For example, Shepard, quotes Einstein’s prologue to Planck’s book (Planck, 1933) as saying, “There is no logical way to the discovery of these elemental laws. There is only the way of intuition, which is helped by a feeling for the order lying behind the appearance.” Taken together with the tacit assumption that thoughts must be either in words or in pictures (and perhaps also the assumption that any “logical” thoughts must be in words), such reports lead to the conclusion that the thoughts that these famous people had were expressed in mental pictures and that it is these sorts of pictorial thoughts that are the real source of their creativity (I will take up the question of the relation between mental imagery and creativity in section 8.5 below).

Of course one also allows for the possibility that a person might think in terms of acoustical or tactile or other sensations, even though these may be less central for most people. But what about the possibility that the form of thoughts is not only unconscious, but is something that never could be made conscious. Many people have assumed that such an idea is pretty nearly incoherent (see, for example, Searle, 1990). Notwithstanding the widespread influence of Freud’s idea of the unconscious, it is still generally assumed that thoughts are the sorts of things that, although they might occasionally slip by unconsciously, nonetheless *could* in principle be made conscious, and moreover if they were conscious they would be experienced as something we hear or see or otherwise perceive. Indeed, Western philosophy has generally made the awareness of one’s thoughts the basis upon which one ultimately justifies the ascription of particular contents to them: You know what your thoughts are about because they are *your* thoughts and you have special privileged access to them through your conscious awareness. If that were not the case, the argument goes, there would be no basis for ascribing one content as opposed to some other content to thoughts (for a discussion of why we need to appeal to contents at all, and for a different view of how we might ascribe content, see Pylyshyn, 1984a).

I will not quarrel with the idea that we have privileged access to the content of our thoughts, nor will I even try to define what it means for thoughts to be conscious. Nevertheless, I claim that there is every reason to believe that (a) the form of one’s thoughts is something that has to be inferred indirectly, just as one infers the form of matter in physics and chemistry, without prejudging the scientific issue based on how we experience our own thoughts, and (b) what one is aware of – the form and content of one’s conscious thoughts – cannot be what plays the causal story in reasoning. In other words, what we are aware of, such as the inner dialogue of one’s thoughts or the pictures one sees in one’s “mind’s eye” cannot be what is responsible for people having the thoughts they have, and for making the inferences they make. The reason for taking this view is that the content of one’s experience is demonstrably insufficient to encompass the content of our thoughts. The sorts of things of which one is aware, such as the words or sentences of the “inner dialogue” or the “mental pictures” one imagines, greatly underdetermine what one is thinking at the time one has those experiences. Consequently something else must be going on, other than more words or more images. Or so I will argue in the remainder of this section 8.2.

8.2.2 Do we think in words?

Consider an example where one appears to be thinking in words. As I type these sentences I think to myself, “I’d better hurry and finish this section or I will be late for my meeting.” Now this is a pretty innocuous thought with little apparent hidden meaning. But look at it more closely. If I “said” that sentence to myself in my inner dialogue, I meant something far more than what appears in the sequence of words. I knew which particular text on my computer screen I meant when I thought “this section”, I knew how much time would have to pass (roughly) in order for it to count as being “late” for a meeting, I knew which meeting I had in mind when I only thought “my meeting,” and I knew what counts as “hurrying” when typing a section of text, as opposed to running a race. And for that matter I knew what “I” referred to, although the sentence I imagined thinking did specify who that was (“I” refers to different people at different times). In

fact sentences *never* say all that their speakers mean. The sentences of your inner dialogue follow such Gricean Maxims (Grice, 1975) as “make your statement as informative as required but not more informative than required” (in other words, don’t express what you think your hearer already knows). More generally, a statement in a discourse is assumed by all parties to be relevant and to be as informative as appropriate. And sentences follow such maxims just as dependably in inner dialogue as they do in external conversation. But if the sentences do not express all that I know or intend, then what I know or intend must take some other form – a form of which I have no conscious awareness. It is no help to say, as one might in discussing an actual overt conversation, that the hearer of the sentence infers the missing parts, because in inner dialogue the speaker and hearer are the same. And if the speaker knows something that remains unexpressed in the sentences of the inner dialogue it just goes to show that the inner dialogue is not doing the work one assumed it was doing. The imagined inner dialogue leaves many unstated things to the imagination of the inner speaker and hearer whereas, as Steven Pinker puts it (Pinker, 1997, p70), “the ‘language of thought’ in which knowledge is couched can leave nothing to the imagination, because it *is* the imagination” – if we think in language then there is nowhere else for the unexpressed parts of the thought to hide.

The problem of expressing the entire content of your thoughts in sentences is even deeper than might appear from this discussion. It was already known in the early history of logic that sentences are too ambiguous to serve as the vehicles of reasoning. For this reason, logicians had to devise more precise formalisms in order to express in a different way the distinct meanings that a sentence could have. So they invented mathematical systems of symbolic logic, such as the predicate calculus. For example sentences like “Every man loves a woman” can express at least two distinct senses (one in which for each man there is some woman that he loves, and the other in which there is a woman such that every man loves her) thus making it necessary to introduce such syntactic mechanisms as quantifiers and brackets in order to express these distinct meanings.

But the problem with the idea that natural language serves as the vehicle of thought, is deeper than the problem posed by the existence of syntactic ambiguities or by the importance of unspoken context in determining the meanings of sentences in inner a dialogue. Words and phrases appear to cut the world up more coarsely than does thought. There are many concepts for which there is no corresponding a word (though presumably for many of them there *could* be an appropriate word). But, even more seriously, one can have thoughts when one is perceiving something, whose contents *cannot* be expressed in words, not even for one’s own private purposes. The outstanding case of such thoughts is one that we have already encountered in chapter 5: they are thoughts with a demonstrative (or indexical) component. I can, for example, think a thought such as “This pencil is yellow” where I am able to pick out an individual and claim of it that it is a pencil and that it is yellow. The object that my visual indexing system picks out is the very object about which I am having the thought and of which I am predicating certain properties. The resulting predication forms part of the content of my thought, yet it cannot be expressed, linguistically or otherwise. And for that matter, neither can most of the properties of the object I am seeing and which can enter into my thoughts. For example, I can have the perfectly clear and well-formed thought “*This* is the *same* color as *that*” where the content of my thoughts in some important sense includes the color, as well as the proposition that the two *things* I have mentally picked out are the *same* color. Yet I cannot express this thought in language – not because I don’t know the correct words, but because I needn’t even have a category or concept for an important part of what I am thinking, namely what is being referred to by the demonstrative “that,” as well as its particular color! Some thoughts, in other words, can contain *unconceptualized contents*. This, in turn, means that the grain of thoughts, or the possible distinctions among their contents, is even finer than those of one’s *potential* linguistic vocabulary (assuming that one’s vocabulary is necessarily confined to the things for which one has concepts).

The idea that natural language cannot be the medium of thought because of inherent ambiguity and instability, in relation to the specificity of the contents that it expresses, has been noted by many other writers.

(Block, 2001) has recently argued that there is also an inherent difference in grain between thoughts and experiences – we can have visual experiences at a finer grain than the grain of our thoughts. (Sperber & Wilson, 1998) make a similar point. They argue that there must be many more concepts than there are words and that, unlike concepts, the meaning of a particular word token may depend on many pragmatic factors. (Fodor, 2001) also makes the same point in arguing that language only approximates the compositionality and systematicity required of thought (see also the discussion of systematicity and compositionality in section 8.2.4).

8.2.3 Do we think in pictures?

Suppose we concede that sensory contents can be finer grained than verbal ones and thus that we can represent properties visually that we cannot describe verbally. This suggests that mental states must include sensory contents as well as verbal contents, which is the basic premise of the dual code view of mental representations, as well as the “perceptual symbol system” idea championed by (Barsalou, 1999). In the case of visual properties, we might postulate that these could take the form of visual mental images. But visual images are just as inadequate in their expressive power as are sentences. Just as the idea that we think in sentences is inadequate because the bulk of the information remains elsewhere, so mental pictures face the same indeterminacy problem. Consider a simple example in which a visual image allows one to solve a certain problem.

People find it easy to solve what are called three-term series problems by imagining that the objects described in the problem are located in an array where the spatial relations among them represents their relative standing on some measure. A typical problem goes like this. “John is taller than Mary but shorter than Susan. Who is tallest (/shortest)?” (Instead of “taller” one can substitute “smarter” “richer” and many other such relations.) To solve this problem, all you have to do is place a tag representing John at a fixed location in an image, then when you hear that John is taller than Mary you locate a tag for Mary below the one for John in a vertical stack, and when you hear that John is shorter than Susan you locate a tag for Susan above the one for John to yield an image such as Figure 8-1. You solve the “who is shortest” problem by finding the lowest tag in the image (and similarly for answering “tallest” or “all who are taller than” questions).

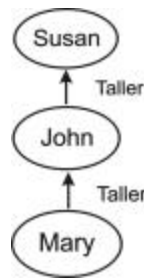


Figure 8-1: Elements in the three-term series problem may be imagined as located in a vertical array and the answers to the problem may be obtained by merely examining the array.

(De Soto, London, & Handel, 1965) refer to this way of solving problems as using “spatial meta-logic” (see also Huttenlocher, 1968). Even though this example is extremely simple it has inspired a great deal of research. For example, the exact wording has been shown to have a strong effect, which suggests that linguistic processes – translating from sentences to some other form of representation – may account for why problems worded in one way are easier than ones worded in another (Clark, 1969). But even here one might ask whether other sorts of cognitive representations might be involved, besides those in the image and/or the sentences experienced by the person solving this problem. A moment’s thought will reveal that there is much more going on than just inspecting the image. Right at the start, in hearing the sentence “John is taller than Mary but shorter than Susan” the hearer must figure out what the missing parts of the sentences

are: the subject and verb of the second clause are missing and must be restored in the course of the grammatical analysis of the sentence. In fact by some analyses even more words are missing in the surface structure (“John is taller than Mary but shorter than Susan” means “John is taller than Mary <is tall> but <John> <is> shorter than Susan <is short>”). This involves appealing to grammatical rules (in this case, deletion rules) – the processing of which is never conscious. Following this grammatical analysis there is still the question of how the problem statement is converted into “mental drawing” instructions. Notice that in order to know that a vertical array is an appropriate structure to use in this problem, the person must already have done some reasoning. Only relationships that have the formal property known as *transitivity* are candidates for such a structure. If the problem had been about the relation “likes” then the structure would have been inappropriate (since “Mary likes John” and “John likes Susan” does not entail that “Mary likes Susan”). How do subjects solve all these preliminary problems? Through what *conscious* imaginal or verbal steps do they go? Obviously they do not do it in language since this is about how they understand the language, and they don’t do it in pictures since it is about deciding what pictures to construct and how to interpret them.

There are also many properties of the imaged array that are assumed in this way of characterizing the problem-solving process. Where do these assumptions come from and how are they justified? For example, according to the account given, a symbol for Mary (call it *Mary*) is placed below *John* in a vertical stack. Next, *Susan* is placed above *John*. This appears also to locate *Susan* above *Mary* in the stack. Why? For that to be true two further things must be true of the image. One is that it must be the case that some operations on the image do not change certain relations that are already there. In the case of an array written on a board, when you add an item above a stack of two items already there, the two items generally keep the same relations as they had before the addition. That is so at least partly because the board is rigid and some geometrical relations on a rigid board remain fixed when new objects and relations are added. But even this is not always the case. For example a graphical representation of the two-placed relations “directly above” or “next to” are changed when another object is inserted between them. Thus if the instructions had led you to place *Mary* “directly below” *John* and the next instruction also led you to place *Susan* “directly below” *John*, the relation between *John* and *Mary* would be changed (or become indeterminate) as a result of the new addition (*Mary* may no longer be “directly below” *John*). (For more examples, see section 8.4).

While this seems like a trivial matter and easily corrected, it turns out to be the tip of a very large iceberg called the “Frame Problem” (see, e.g., Ford & Pylyshyn, 1996; Pylyshyn, 1987). The problem arises because when one carries out actions in one’s mind, as opposed to actually executing actions on the world, the problem of determining from what one knows, what will change and what will remain fixed is in general intractable. That’s because what can change is whatever may be relevant to the action, and the general problem of relevance is itself intractable; any belief can be relevant to any action so long as there is a possible chain of reasoning that could connect them. Such a chain always exists since one can in principle have beliefs that conjoin any pair of other beliefs. One could have the belief “if P_j then P_k ” for any possible pair of beliefs P_j and P_k , and so if P_a is relevant to action a , then there may a chain from any belief P_j to action a . For example, supposing that you have the belief (P_1) that to leave the room (i.e., to be located outside the room) you should open the door, and that to open the door you must be close to the door (P_2). If you also have the apparently unrelated belief (apparently unrelated because it does not involve any of the concepts of the first belief, such as door, inside, etc) that there is a security system triggered by movements that locks all means of egress (P_3), then the link between P_1 and the goal of being outside the room is blocked. This information may be implicit in your knowledge, but inferring it could mean checking every belief in your entire database of beliefs – an intractable task given the nearly unbounded size of your entire belief set (for more on this “Frame Problem” see the essays in Pylyshyn, 1987).

Even if you could correctly solve the problem of whether an already-present relation in the image changes or remains unchanged when an action (including checking the image for the presence of some property) is

performed, there is then the problem of interpreting the resulting image. Consider Figure 8-1 to be an exact rendering of the image from which you reason this problem through (notwithstanding the argument in section 6.5.5 that this can't be done in principal). Why does the shape and size of the ellipses or the size and font of the names not enter into the conclusion? What determines which relations are relevant and how to read them off the image? Also in this example the same meaningful relation ("taller") occurs in different geometrical guises. While the diagram shows John being taller than Mary and Susan being taller than John, it does not show that Susan is taller than Mary in terms of the same geometrical relationship (e.g., the symbols for Mary and Susan are further apart and there is an intervening item in the array). Reading off this relationship requires knowing that the presence of an intermediate item does not affect the relationship between the two items in question (unlike the relationship "adjacent to"). This is a property that is not represented in the image or anywhere else in consciousness. In each of these examples, reasoning is involved that is not represented in the diagram, nor is it contained in an inner dialogue. The point of this mundane analysis is that the things of which we are aware – the words and mental images – are never sufficient for the function attributed to them. In every case, more is going on, as it were, "offstage". And what more is going on is, at least in certain cases, patently not expressible as a sentence or a picture. Yet despite this important point, it is nonetheless also true that pictures can enhance thinking in important ways that I will take up in section 8.3, because vision, when directed to real diagrams, often provides operations that can be used to draw inferences more simply than by using rules of logic. This is especially true when reasoning about spatial properties, as in geometry.

8.2.4 What form must thoughts have?

(Fodor, 1975) presents a tightly argued case that thought must be encoded in what he calls a "language of thought" or LOT. Of course LOT cannot be a natural language, such as English for the reasons presented above. Also, of course, a great many organisms do not have any language, as is the case with pre-linguistic children and non-linguistic organisms like higher primates, and yet are clearly capable of thought. Any vehicle of thought must have certain properties. One of these properties is *productivity*. Since the number of thoughts we can have is, in principle, unbounded (except for practical considerations such as one's mortality and boredom), thoughts must be constructed from simple elements, called concepts (as Humboldt put it, our representational capacity is capable of "infinite use of finite means" ²⁶). Moreover, the meaning of complex thoughts must be derived from the meaning of its constituent parts, which means that thoughts must be *compositional*; all the meaning of a complex thought must come from the meaning of its canonically distinguishable parts, together with the rules of composition (or syntax) of LOT, it cannot come from any other source. Even more important, thoughts must have the property that (Fodor & Pylyshyn, 1988) call *systematicity*. What this means is that if an organism is capable of thought at all it must be capable of having a set of related thoughts. If it can think, say, "Mary hit the ball" and it can think "John baked a cake" then it can also think the thoughts "John hit the ball" and "Mary baked a cake" and even the thought "John hit the cake" and "Mary baked the ball". Of course it is unlikely to think the latter because these are bizarre events, but the point is that the system for encoding thoughts encodes complexes from simples and once that system is in place and the constituent concepts are available, the capacity to form the complex thoughts is inherently there.

In (Fodor & Pylyshyn, 1988) we also argued that inference requires compositionality of the system of representation. Rules of inference are schemas; they tell you how to draw a conclusion from premises by recognizing the constituents of a structured representation and transforming them in ways that preserve truth (or some other semantic property, such as plausibility). For example, if you have a thought with the content,

²⁶ This idea was made famous in the early 1960s by Noam Chomsky. For more on the Chomskian revolution in Cognitive Science see: <http://mitpress2.mit.edu/e-books/chomsky>

“It is dark and cold and raining” you are entitled to conclude, “It is raining” (or “it is cold” or “it is dark”). This follows from the rule of conjunction elimination: from “P & Q” infer “P” (or infer “Q”). In this rule the “P” and “Q” can be replaced by arbitrarily complex expressions. For example, in a particular case the position occupied by “Q” might instead be occupied by “R and (S or T)” in which case one would be entitled to infer from the expression “P and (R and (S or T))” that “P” or that “R and (S or T)”. Take a thought such as the one given earlier (“it is cold and raining”). Since any intelligent creature could infer from the thought “it is cold and raining” that “it is raining” it must also be able to infer from “it is dark and cold and raining” that “it is raining”. In other words there is systematicity in what you are able to infer from complex thoughts: If you can make one sort of inference, you are thereby also able to make other sorts of inferences (especially simpler ones as in the example here). None of this depends on exactly how the beliefs are encoded except that they must have recognizable parts that determine the meaning of the whole and these parts, in turn, are what determine what can be inferred. These sorts of considerations are what lead one to the view that whatever form thoughts may take they are bound to meet the conditions of systematicity, from which it follows that they are bound to have constituents and to be compositional. This makes them rather more like a language than a picture or any other fused form that does not recognize semantically interpreted parts, such as proposed by the “connectionist” approach to modeling mind (for a more detailed exposition of this point see Fodor & Pylyshyn, 1988).

Notwithstanding the argument presented above, the requirement of compositionality does not prevent there being analog components in representations. For example, there is no prohibition against individual terms in the vocabulary of LOT being drawn from an infinite set, as they might be if they constituted what (Goodman, 1968) calls a “dense symbol system.” While such arcane symbol systems are not prohibited, it is still an open question what would be gained from assuming them. We know, more or less, what can be done with discrete symbol systems (pretty much all of logic and computer science is about such systems), but there are only hints as to what would be gained by admitting other sorts of symbols into LOT. One can also augment LOT by adding non-linguistic (or non-conceptual) elements in several ways. For example, there is a different way in which nonlinguistic entities can function in thought; they can provide external elements or models that can be used in reasoning. We are all familiar with the usefulness of diagrams, charts, graphs, and other visual aids in conveying ideas. Before considering the possibility that parts of thought may be carried by non-symbolic means, we need to look at why real diagrams might be useful in reasoning and ask whether those functions can be reasonably obtained from imagined diagrams.

8.3 How can visual displays help us to reason?

8.3.1 Diagrams as logical systems that exploit visual operations

To accept the conditions on the format of reasoning is not to deny that thinking about some kinds of problems can be enhanced by the use of vision – real vision of real displays. Some kinds of problem solving – say proving theorems in plane geometry – appear to proceed more efficiently when we are able to use displays or diagrams as a prosthetic to aid in thinking. Although the mathematician David Hilbert showed that Euclidean geometry using diagrams and based on Euclid’s axioms (as put forward in his “Elements”) is not rigorous in that it makes hidden assumptions (particularly about the existence of points and lines that are drawn according to the compass-and-straightedge requirement), it is difficult to imagine proving a theorem in plane geometry without drawing a figure. But consider the assumptions that go into some of the simplest proofs from Euclid. To prove his First Proposition, Euclid used one diagram and accompanying text. The proposition says that an equilateral triangle can be constructed on a given line AB using only compass and straightedge. The construction is obvious from the diagram: Draw a circle with one end of the line (A) as center and then another circle centered on the other end of the line (B). Then join the point where they intersect to each end of the line. It is an equilateral triangle by construction. The trouble with this “proof,” as

many people have remarked, is that it assumes without proof that the two circles *will* intersect – an assumption that does not follow from any of Euclid’s “elements” or axioms.

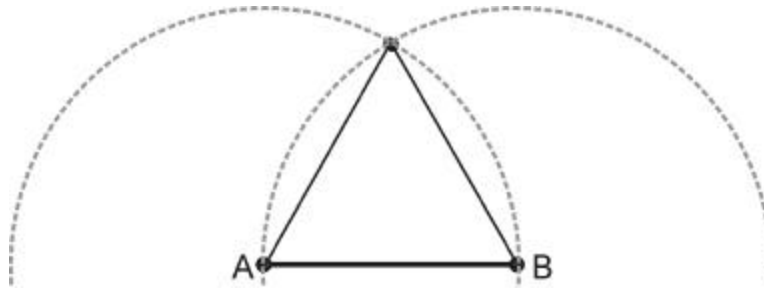


Figure 8-2. Euclid’s First Proposition: an equilateral triangle can be constructed on a given line using compass and straightedge.

The general consensus had been that diagrams could be (or even must be) dispensed with in order to provide rigorous proofs of Euclidean theorems. Hilbert provided an axiomatization of Euclidean geometry that had a very non-figural character and did not use geometrical constructions; using these axioms Hilbert was able to make Euclid’s propositions rigorous and free from hidden assumptions. The question of whether diagrams are merely heuristic or whether they can provide a rigorous means of proving theorems has been controversial over the years. But the question has received renewed interest in recent years. A number of writers (e.g., Allwein & Barwise, 1996) have argued that diagrammatic reasoning can be made rigorous while retaining its diagrammatic character. Recently (Miller, 2001) developed an axiomatic system (EG) that treats elements of diagrams as syntactic objects and proves theorems about them that are nearly as simple and transparent as conventional Euclidean proofs. The system has even been implemented as a computer program (called CDEG – for Computerized Diagrammatic Euclidean Geometry). While this system reasons with syntactic objects that correspond to parts of diagrams, it does not conjecture possibilities based on appearances, thus missing one of the important aspects of human diagrammatic reasoning. Miller was interested in providing a sound formal system, as simple as that based on Euclidean postulates, to back up the informal use of diagrams to explore geometrical relationships, and in this he was eminently successful. Our goal in this section, however, is to examine how diagrams and the human visual system together could form a more powerful system for reasoning than unaided cognition (or perhaps than reasoning aided by a notepad for writing down formulae or for keeping track of cases).

The visual system is one of our most developed and exquisite cognitive faculties and we make use of it in many more ways than recognizing familiar objects and in guiding our movements. For example, we are able to convert many abstract problems into spatial form, as we do when we reason using Venn diagrams, graphs and other modes of spatial representation (including the example illustrated in Figure 8-1). Venn diagrams (which are one of the more successful of a wide range of logic-encoding diagrams discussed by Gardner, 1982), allow one not only to illustrate relations among sets or among propositions (the two are interdefinable) but also allow one actually to use the diagram for logical reasoning. The Venn diagram representation of propositions, an example of which is shown in Figure 8-3, illustrates that certain properties of the visual system (e.g., the ease with which it detects whether one region overlaps another or whether some designated element is inside a particular region) can be exploited to facilitate reasoning. Such externalizations exploit the perceptual system (usually, though not necessarily vision) to help recognize patterns. In accepting that vision can play this role we need make no assumptions about the form of our thoughts, only about how we are able to map from perceived patterns to thoughts in the pursuit of problem-solving goals. Thinking while seeing allows our thoughts to exploit important properties of space without assuming that the form of our thought is itself spatial.

Venn diagrams and other graphical forms of logical inference are more than just mental crutches; they allow rigorous proofs to be formulated, as (Allwein & Barwise, 1996; Jamnek, 2001) have shown. But for this to occur, the graphical form, together with the appropriate operations and interpretations placed on them, must be isomorphic to normative rules of logic. The study of such graphical forms has been a serious pursuit since at least the time of Ramon Lull, an early 13th century scholar, who invented an influential method, called *Ars Magna*, which consisted of a set of geometrical figures and mechanisms for guiding reasoning. (Gardner, 1982) provides a fascinating discussion of such devices, which include the diagrammatic inventions of Leonhard Euler, Sir William Hamilton, Allan Marquand, Johann Lambert, Charles Peirce, Lewis Carroll (Charles Dodgson) and Gerrit Mariè Mes. But diagrams and other sorts of figures can also be useful even when they are not embodied in a rigorous system of valid reasoning. Indeed, the use of diagrams in proofs of plane geometry is of this non-rigorous sort. So the question remains, how and why can diagrams be useful as heuristics in such cases? Clearly there is something about vision that provides functions that are not as readily available in other forms of reasoning. And if that is so, then perhaps mental imagery provides similar functions. It is worth considering, therefore, what vision might contribute to reasoning and perhaps also to creative reasoning.

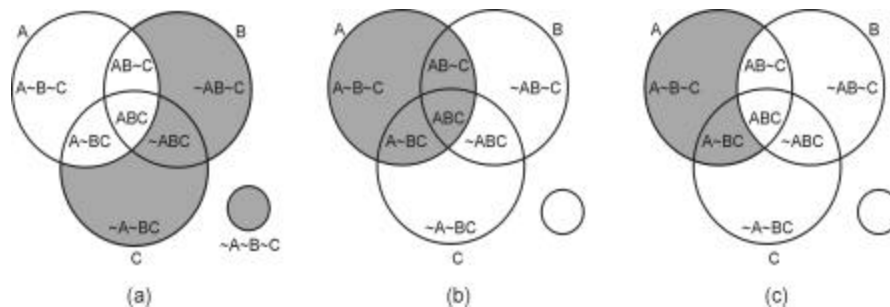


Figure 8-3. Venn diagram representing the propositions A , B , C , their negations $\sim A$, $\sim B$, $\sim C$ and their pairwise combinations. Figure (a) shows A as true, which leaves everything that contains $\sim A$ false and shown as shaded. Figure (b) shows A as false (and therefore shaded), and leaves all the other possibilities true. Figure (c) shows how material implication (denoted $A \rightarrow B$), can be represented, by shading all regions in which A is true and B is false (all regions containing $A\sim B$). To see what is entailed by $A \rightarrow B$ when A is false, shade all areas of (c) containing A that are not already shaded; then we get figure (b) and we see that B can be either true or false. To see what is entailed by A being true, shade all areas in (c) not already shaded that contain $\sim A$. This leaves only the two central regions ($AB\sim C$ and ABC) unshaded and leaves all areas containing $\sim B$ as shaded. So if $A \rightarrow B$ then A being true entails that B is true. (Based on Gardner, 1982).

8.3.2 Diagrams as guides for derivational milestones (lemmas)

Consider the following example in which drawing the right diagram makes it easy and transparent to prove one of the most important theorems in plane geometry: Pythagoras's Theorem. Figure 8-4 shows a right angle triangle (shaded) with sides a , b , and hypotenuse c . To prove that the square on c is the sum of squares on a and b , we begin by drawing a square on side c . Then we extend the sides a and b until they both equal $a + b$ and we draw a square on these sides (the area of this square is $a + b$ times $a + b$). Without going into the individual steps one can readily prove that the original triangle is reduplicated 4 times inside this large square (by checking the angles, verify that each of the triangles in the corners is similar to the original triangle and because each has hypotenuse of length c , they are congruent to the original triangle). Thus we see that the large square, whose area is $(a + b)^2$ is made up of the square on c plus 4 triangles. Therefore to get the area of the square on c we subtract the 4 copies of the original triangle that fit between the square on c and the outer square on $a + b$. Since the area of each of those 4 triangles is $\frac{1}{2}ab$, we have the following equation: $c^2 = (a + b)^2 - (4 \times \frac{1}{2}ab) = a^2 + b^2 + 2ab - 2ab = a^2 + b^2$.

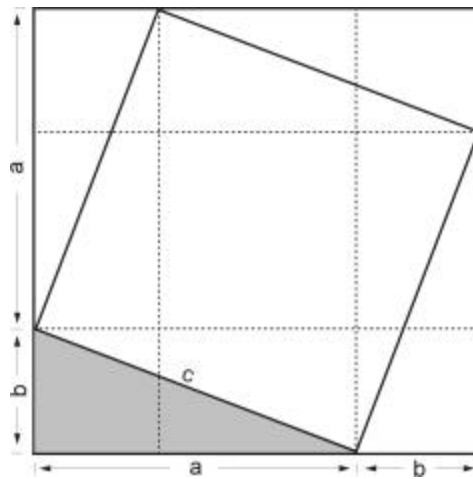


Figure 8-4. Illustration of how construction helps to prove a theorem in plane geometry, in this case the important Theorem of Pythagoras.

This proof, though easier to describe, is not as elegant as some of others that do not use algebra²⁷ (there are at least 367 published proofs, many of which only require folding and shearing and overlaying parts of the constructed figures, see the references in Dunham, 1994), but it does illustrate the importance of construction in geometrical proofs. Without drawing the square on c and on $a + b$ the proof would be extremely difficult and in fact would likely have been impossible in Euclid's time. So what purpose do the constructions serve? In Euclid's geometry the constraints on construction defined the rules of plane geometry: all problems could be paraphrased in the form, "using only a compass and straightedge, show that ...". The constraints placed on construction constituted one of Euclid's 5 "postulates." But beyond that, selecting which intermediate figures to construct is like deciding on which lemmas to prove in the course of proving the main theorem. In

²⁷ A selection of proofs of the Pythagorean Theorem, many of them animated and interactive, can be viewed at <http://www.cut-the-knot.com/pythagoras/>

addition, the visual detection of similarity and congruence (as we saw in Figure 8-4) provides the important guidance in conjecturing lemmas and developing the proof.²⁸

The way in which diagrams and other visual aids help us to formulate and solve problems is far from being understood. The following characteristics of vision are surely important to that pursuit:

(1) Vision provides primitive operations for a number of functions, such as shape recognition and the detection of relational properties like the “inclusion” of certain regions within other regions (in fact this particular relational property formed the basic operation out of which Jean Nicod hoped to build his “sensory geometry,” see Nicod, 1970). Primitive detection operations such as these are very important, for example, in exploiting Venn diagrams. The usefulness of diagrams, graphs, charts, and other visual devices, relies on the fact that people are very good at visually detecting certain geometrical relations (as we saw in the discussion of “visual routines” in Chapter 5).

(2) Another important property of vision, which has received considerable attention in recent years, is that it appears to use a strategy of keeping track of where information is located in the world, rather than encoding it all in memory, thus in effect providing what amounts of a very large memory. The idea that vision uses “the external world as memory” was argued forcefully by (O'Regan, 1992) and supported by findings of (Ballard, Hayhoe, Pook, & Rao, 1997). Ballard et al. showed that in carrying out a simple task such as constructing a copy of a pattern of colored blocks, observers encoded very little in each glance (essentially only information about one block), preferring instead to return their gaze over and over to the model figure that they were copying.

In chapter 5 I argued for the general importance of keeping track of information in the world that had not yet been encoded (or conceptualized), and proposed the mechanism of visual indexes (or FINSTs) to allow gaze or focal attention to return to certain parts of a figure whose properties had yet to be encoded. Thus although there is now considerable evidence that surprisingly little of a scene is recalled from a single glance (Rensink, O'Regan, & Clark, 2000; Simons & Levin, 1997) – unless the observer happens to fixate or attend to a particular part of the scene (Henderson & Hollingworth, 1999; Rensink et al., 1997) – yet a great deal of information is potentially (as well as phenomenologically) available. (To anticipate the discussion in the next section I might note here that the mechanism of visual indexing can only be exploited when perceiving a real scene, not when accessing information from a mental image).

(3) Vision routinely goes beyond the information given – a percept is invariably a generalization of the individual properties of a unique stimulus. In fact any recognition of a pattern is just such a generalization: We do not see something as a chair except that we generalize the particular shape in front of us by putting it in the same category as an unlimited number of other shapes, namely all the shapes we see *as* chairs. The naturalness of such visual generalization is an essential part of reasoning with diagrams and will be discussed in the next section.

8.3.3 Diagrams as a way of exploiting visual generalization

When we draw diagrams and examine them to determine whether some general property holds, we do not just rely on whether we detect an instance of that property in the particular diagram; we often appear to “see” what properties will *necessarily* hold of any resulting construction. For example, if you want to know

²⁸ Barwise & Etchemendy (1996) use the same diagram but view its construction differently. They start off with the original triangle and construct a square on its hypotenuse. Then they explicitly replicate the triangle three times and move it into the positions shown in Figure 8-4. The proof then proceeds by showing that these triangles fit snugly in place by using the fact that three angles of a triangle sum to a straight line. Replicating and moving/rotating a triangle are operations that would have little meaning without a diagram.

where the intersection of two lines will be located if they are drawn inside a rectangle according to a certain specification, you have but to draw an *arbitrary* rectangle and then draw *arbitrary* lines in it meeting the specifications and you can quite often see not only the location of the intersection of particular pair of lines that you drew, but also where the intersection must fall for *any* pairs of lines in *any* rectangle that meet the given specifications. For example, you can tell that if you draw a line from each of the two bottom vertices of any rectangle to any point on the opposite vertical side, then these two lines will (a) always intersect and (b) the intersection will always lie below the horizontal midline of the rectangle. You can *see* that by just looking at a particular instance, as in Figure 8-5. How you turn a particular instance of a figure into a universal generalization is far from obvious (although one can certainly come up with plausible hypotheses for particular cases). While going from a particular to a universal is no doubt related to how one does abductive reasoning in general, it appears that in this case the visual system is involved essentially. Its involvement, moreover, goes beyond merely recognizing that a certain particular pattern or property is present in a particular instance of the drawing. Visual perception appears to be the source of the generalization in the first instance.

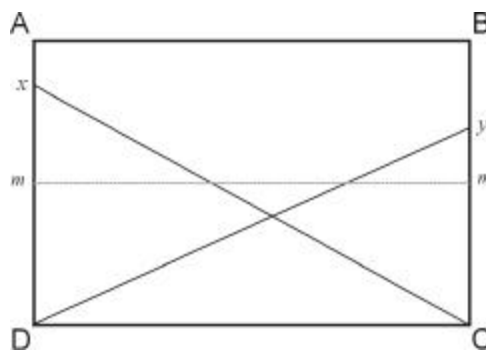


Figure 8-5. You can “see” from this one figure that if lines (Dy, Cx) are drawn from the bottom vertices to any point on the opposite side, the lines will meet at or below the midline (m-m') of the rectangle for any possible lines drawn to specification and in any possible rectangle.

How does vision make the jump from a particular to a general? One could visually encode general properties by such devices as annotations added to a diagram in various ways to show some general property. Sometimes these annotations are diagrammatic and so do not appear to be annotations. For example, we could represent the infinite family of lines implied by the above drawing in some graphical manner (say as a fan-shaped region, such as in Figure 8-6). This would still be a particular token shape, except that we would know to interpret the shaded area as a set of possible lines, much as we would interpret a label. We would thus be relying on an interpretation (which might be quite natural, as in this case) to represent a quantified proposition such as “all lines lying in the region bounded by ...” In this case the representation contains more information (it contains the limits of possible line locations) and would in part be “read” like a descriptive annotation.

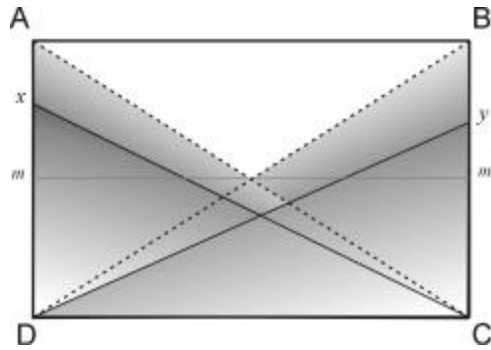


Figure 8-6. One way to think of the class of all lines from D to an arbitrary point on the opposite side is to imagine a fan-shaped area that could be occupied by any such line. Such an apparently pictorial way of representing a class is no different from providing an annotation (perhaps with arrows pointing to particular parts of the diagram) that would be interpreted by the viewer, much the way a symbol would be. In this case, however, the annotation has to refer to sets of locations within the figure. The diagrammatic form of the annotation is particularly useful because it provides a natural way to refer to particular visual objects and locations.

The use of this sort of implicit annotation (where visual properties are used not to depict a visual or spatial property, but to bias the way spatial properties are interpreted by the viewer) is quite common in graphs and is often the only way that noncommittal or vague information can be depicted graphically (see note 7). It is also one of the techniques adopted in many practical schemes for graphing (or “visualizing”) information. Such schemes place a heavy emphasis on selecting the most natural and perspicuous visual properties for conveying various types of secondary or annotation information, such as the use of certain visual dimensions for conveying magnitudes and relationships among magnitudes (this has been superbly exploited by Tufte, 1990, in his examples of visual communication). For example, practical schemes consider a variety of image properties, from physical size to brightness and color, as potential dimensions to convey both magnitudes and their location (e.g., on a map). (Norman, 1988, 1991) illustrates how things can go wrong if relative magnitudes are depicted on a map using a non-metrical visual property, such as texture.

But it may be misleading to view the process by which diagrams depict general properties as equivalent to adding annotations. The process of going from a particular to a universal is inherent in how vision works; how it parses a complex figure into subparts and how it represents and recognizes patterns. For example, in section 3.1.1 we saw that perception of 3D objects makes essential use of the distinction between accidental and nonaccidental properties. Accidental properties are ones that depend on the precise viewpoint and disappear when the viewer makes the smallest movement. For example, two lines in an image could appear aligned or could appear to be coterminous (connected) entirely by accident. They could be two randomly situated lines in 3D that are seen from a particular point of view which just happens accidentally to align them or causes their endpoints to be at the same place on the 2D image (as illustrated in Figure 3-3 of Chapter 3). Yet the probability of this happening by accident (i.e., the probability of the 2D image containing two linearly aligned or coterminous line segments when the lines in the world were actually not aligned or coterminous) is vanishingly small and could be detected by a slight change in viewpoint. Consequently the visual system always interprets such occurrences as non-accidental (the principles by which the visual system provides nonaccidental interpretations of image features are discussed in Chapter 3, especially section 3.1.1.2). In making such an interpretation, the visual system appears to jump to a conclusion about a universal property from a single instance. Such apparent abductive leaps are quite characteristic of vision and may form the basis for how vision helps us to reason: It provides non-demonstrative inferences on matters pertaining to shape (especially 3D shape which has been studied most).

While it is not viewed in exactly these terms, the study of computational vision – and in particular, the study of object recognition and representation – centers precisely on the problem of how vision generalizes from particulars, or token image shapes, to classes of image shapes. When we “recognize an object,” what we do is map some of the geometrical properties of the token proximal image onto properties that the object *class* possesses. That’s because “recognizing an object” means, in part, to assign the image to a category, and in part it is to provide a representation of the shape that all members of the category possess. I said “in part” because this is only part of the story. While in a sense vision strips off the properties that are specific to the image token in order to focus on the properties of the general object class, vision generally does not throw this information away. It usually represents some of these particulars separately, for example as parameters. This is what various schemes for object representation do; they allow us to factor apart various aspects of the object shape. For example, it may factor out the object’s size and orientation (using some natural axis) and then represent these as separate aspects of the object’s shape. It may also encode the way in which the shape deviates from a simple core pattern. Several object-perception methods use what are called “deformable models” to represent a general shape. One common method is based on wrapping a mathematically simple skin over the shape, using a formalism called a “superquadric” (Pentland, 1986). Many such schemes for representing object shapes have been used in robot vision. One of the most widely adopted is the generalized cylinder (or generalized cone) representation, initially proposed by (Binford, 1971) and developed further by (Marr, 1982). This representation characterizes 3D objects in terms of structures of cylinders whose primary axes can be curved and whose radii can vary across the length of the cylinder. In other words, every shape is described as some set of interconnected and deformed cylinders. This form of representation has been adopted as part of a theory of object-recognition known as Recognition-by-parts, developed by Irving Biederman (Biederman, 1987; Biederman, 1995).

Another interesting example of a core-plus-deformation scheme for object representation was proposed by Michael Leyton (Leyton, 1992). In Leyton’s theory, an object’s core representation is a highly symmetric shape (for example, a sphere), and the transformations represent various ways that this simple core shape might have been transformed in order to yield the observed shape. The transformations in Leyton’s theory are not just a mathematically compact way of representing shapes, but are chosen to reflect various salient ways in which *physical* processes might operate to transform the core shape, hence the representation can be viewed as providing a reconstruction of the object’s causal history of physical deformations. The shape transformations include those produced by physical forces such as stretching, squeezing, twisting, skewing, and so on, as well as other factors such as biological processes like growth. They also reflect optical processes such as those involved in the projection of light from object to image, which result in transformations such as those of perspective.

There are many ways of representing shapes in terms of some canonical shape plus some parameters or shape plus transformations. For our purposes I wish simply to note that choosing a form of representation, such as the one based on the generalized cylinder, establishes a pattern of generalization. For example, in the recognition-by-parts technique (which is based on deformed cylinders), a shape will generalize more readily to objects that have the same relational structure of cylindrical axes than to those that do not, more readily to objects with the same profile of axis diameters than those that do not, and so on. Biederman’s recognition-by-parts system inherently generalizes shapes in just this way, as he showed experimentally by examining patterns of confusion in the recall of shapes (O’Kane, Biederman, Cooper, & Nystrom, 1997). In particular, the system is relatively insensitive to size (Biederman & Cooper, 1992), viewpoint (Biederman & Bar, 1999) and rotation in depth (Biederman & Gerhardstein, 1993). To the extent that the theory is a valid description of what early vision computes, it embodies a scheme of visual generalization that functions when we visually perceive patterns, including diagrams.

8.3.4 Diagrams as ways of tracking instances and alternatives

A diagram used in problem solving is often nothing more than a representation of a particular instance of the givens in a problem; a representation of some token situations, rather than of the entire class of situations mentioned in the problem statement. In some cases diagrams can be useful, even without the special properties of visual generalization, because they encode distinct cases and allow us to keep track of them.

Consider the following example, discussed by (Barwise & Etchemendy, 1990). The problem is stated as follows: “You are to seat 4 people, A, B, C, and D in a row of 5 chairs. A and C are to flank the empty chair. C must be closer to the center than D, who is to sit next to B. Who must be seated in the center chair and who is to be seated on either end?” Here is a straightforward way to solve this problem. First set out the 5 chair positions like this: – – – – –. Next consider the constraints. Since A and C are to flank the empty chair there are 6 cases (where ? indicates the empty chair and – indicates a free or unallocated chair):

Case 1:	A	?	C	–	–	Case 1a:	–	–	C	?	A
Case 2:	–	A	?	C	–	Case 2a:	–	C	?	A	–
Case 3:	–	–	A	?	C	Case 3a:	C	?	A	–	–

Where cases 1a, 2a, 3a are the mirror images of case 1, 2, and 3. Since none of the constraints refer to left or right, the problem can be solved without considering those cases separately, so we ignore the cases 1a, 1b, and 1c. Since C must be closer to the center than D we can eliminate Case 3, which place C further from the center than D. Since D must be next to B we rule out Case 2 because it does not have two adjoining unassigned seats. This leaves Case 1, with B and D assigned as follows:

A	?	C	B	D
A	?	C	D	B

In these remaining cases (or their mirror images) all the stated constraints are satisfied. So the answer to the question is: C must occupy the middle seat and A must occupy one end seat, while either B or D can occupy the other end seat.

While Barwise & Etchemendy present this as a problem that is easily solved with the aid of *vision*, in fact the diagram makes very little use of the many specific properties of vision. The diagram primarily serves the function of setting out the distinct cases and keeping track of the names that occupy the 5 seat locations. Other than the ordering of the 5 names and their easy access by vision, there is nothing “visual” about the method. Of course a visual figure is an excellent way to encode which position has which letter and it does make it easy to “read off” the contents of the 5 locations. But no property of the pattern other than name and ordinal position is relevant. Any way of encoding name information and any way of representing ordinal position would do (e.g., this information could have been represented in the tactile modality or as a melody in the auditory/temporal modality). I do not wish to make too much of the modality in which the figure is presented (vision is most likely the easiest if it is available), but it is important for our purpose to see exactly what function the “diagram” is serving. It is allowing us to construct the distinct cases, to apply the problem constraints to each of them, and to access their contents readily. Because the constraints are stated in terms of the location in a row, a linear array is well suited. And because the visual form makes the array easiest to re-access, it may offer an advantage over audition. The fact that only ordinal information together with the identity of individuals (as given in the problem statement) is relevant also means the problem could easily be solved by a computer using a list representation.

Here is another way of characterizing what is special about such a use of a diagram. It is a constructed instance. In such a pure construction (one without symbolic annotations) there are no quantifiers (like “some,” or “all,” or numerical quantifiers such as “some number greater than 2”), there is no negation (the constructed instance cannot represent that something *may not* be the case), and no disjunctions (they cannot represent that either one case or another holds except by the convention of listing the entire disjunction of

alternatives separately, as I did in the above example, where disjunction is implied. The list format could not directly represent the constraint that the item at the end had to be “A OR D”). Pure constructions contain individuals, not classes of individuals, and properties or relations are ascribed to existing individuals. This appears to be the essence of constructions such as those in this example. To be sure, diagrams have other properties as well (e.g., they represent *metrical* properties, as well as other visual-appearance properties, such as color, shape, orientation, spatial frequency), but I will put these aside for now since I want to consider the function that can be played by *mental diagrams*, where metrical properties present special problems (contrary to some claims about mental imagery). I therefore concentrate on the question of what constructions would be like if we abstracted from these spatial and metrical properties. It turns out that this question has been explored formally (by Levesque, 1986) under the name “vivid representation,” which was discussed briefly in Chapter 7 (section 7.6.2). Such representations have some interesting computational properties that recommend them as a potential formalism for certain kinds of reasoning (what Johnson-Laird, 1989, calls “model-based” reasoning). Operations over such a formalism can be very efficient.

8.3.5 Diagrams as non-metric spatial models and spatial memory

There is little doubt of the importance of spatial constructs in organizing our thoughts, including at the very abstract levels. Spatial metaphors abound when communicating complex ideas and in working through problems (Talmy, 2000). Yet it is far from clear how the mind uses spatial concepts and spatial organizations to help us to communicate and solve abstract problems. This much seems clear: We are able to exploit certain properties of vision in order to reason in some non-visual domains. For example we can exploit such abilities as being able to individuate elements, to parse the world into figure and ground, to rapidly recognize spatial patterns, to map a 2D display onto a 3D shape, to perceive motion from successive displays, to complete partially occluded patterns, and so on. Deductive reasoning can exploit a figure's stability and accessibility in various ways, from those involving the use of logical diagrams (such as Venn diagrams) to simply allowing expressions to be written down and transformed while being visually examined.

Of course we can also use spatial displays or diagrams to reason about spatial layouts, and this may in fact be their major use. If you were to design the layout of a new office, it seems inconceivable that you would do so without making a sketch since both the constraints and the final outcome must be expressed in terms of spatial layouts. Consider the following problem (based on an example of Mani & Johnson-Laird, 1982). A subject is read sentences such as the following and then asked to recall them or to draw conclusions about the relative location of specified objects:

- (1) The spoon is to the left of the knife
- (2) The plate is to the right of the knife
- (3) The fork is in front of the spoon
- (4) The cup is in front of the knife

This description is satisfied by the layout of objects shown in (5):

- (5) spoon knife plate
 fork cup

But a change of one word in sentence (2) making it “The plate is to the right of the spoon” results in the description being satisfied by two different layouts, shown in (6):

- | | | | | | | | | |
|-----|-------|-------|-------|--|----|-------|-------|-------|
| (6) | spoon | knife | plate | | OR | spoon | plate | knife |
| | fork | | cup | | | fork | | cup |

As might be expected, the second set of the sentences results in poorer recognition memory as well as longer times to draw conclusions concerning such questions as where the fork is in relation to the plate. From examples such as these (Johnson-Laird, 1989; Johnson-Laird, 2001) concluded that (a) propositional

reasoning is carried out by constructing models and (b) having to deal with a single model results in better performance than having to deal with several such models. This appears quite plausible (at least for reasoning about such spatial arrangements) when subjects are allowed to sketch the layouts. We can see from the example that in the case of the diagrams on paper, the properties being exploited include the ones discussed earlier: visual recognition of relations like “above” and “to the right of”, the visual capacity to detect relations among disparate objects while ignoring their intrinsic properties (perhaps by using visual indexes), and the capacity of the visual system to generalize away from accidental or secondary properties (like the fact that being to the right of and above something entails being to the right of it). With a real diagram the reasoning proceeds more easily because the diagram provides a rapidly accessible memory that simultaneously contains the given relations among the specified objects, and the relations remain fixed as the objects are examined. But what about the use “mental models” for reasoning where the model layouts are in the mind? But a mental diagram does not enjoy the property of spatial stability, which ensures the consistency of spatial relations detected by vision. So why should it be helpful? I will return to this question below.

8.3.6 Diagrams drawn from memory can allow you to make explicit what you knew implicitly

Closely related to the use of diagrams to construct qualitative models of spatial relations (as in the example above), is the use of spatial displays, together with the operations provided by vision, to carry out inferences concerning spatial properties that depend on quantitative relations. Consider the following example. Suppose you have seen a map that contains five objects (e.g., a beach, a windmill, and a church) located as in Figure 6.4 (a portion of which is reproduced as Figure 8-7 below).

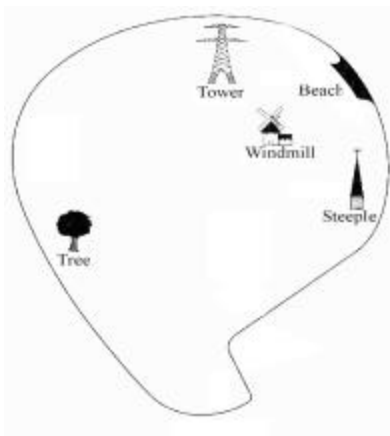


Figure 8-7. Portion of map discussed in Chapter 6, as recalled from memory.

By looking at the map we can see many different spatial relations that relations that hold among the elements and their locations. If I asked you which elements are collinear you can easily provide the answer just by looking. Similarly for questions like which elements are halfway between two others or which ones are at approximately the same latitude, or are furthest south (or east or west) or furthest inland, and so on. We have excellent property detectors in our visual system (I have already discussed the sorts of easily-detectable n -place predicates that rely on what we called “visual routines”). But we have already seen that if you are asked to detect such properties in your mental image you will not do very well on many of them. Your exquisite visual facility does not apply to your mental image.

Now suppose you learn this map to criterion so you are able to reproduce the locations of the 5 distinct features. You may have remembered the locations of these features in any of a very large number of ways: you might for example, have noticed the cluster of 4 features at the top right, or you might have remembered

the places in terms of units of measurement in relation to the width of the island, or you might have recognized a similarity of shape of the island to the African continent and remembered the map features by associating them with some African countries, and so on. Whatever the mnemonic you use, it must of necessity fail to encode an indefinite number of additional relationships that hold among these 5 places, since it is possible to reproduce the location of the 5 objects accurately without having encoded all the binary or triple relations among subsets of objects.²⁹ Thus, even though you memorized the map accurately, you might not have noticed that the tower is at about the same latitude as the beach, or that <tower, windmill, steeple> or <beach, windmill, tree> are collinear or, or that <tower, beech, steeple> or <tower, beech, windmill> or <tower, steeple, tree> form isosceles triangles, and so on. Since there are a very large number of such relationships, you are bound to have failed to explicitly encode many of them. Yet despite failing to encode such relationships explicitly, if the representation allowed an accurate diagram to be drawn, it would be possible to notice these and indefinitely many other such relationships by drawing the map from the information you did encode and then *looking at the resulting drawing*. The missing relationships are implicit in the accurate location of the 5 features. This, in a nutshell, is the advantage of drawing a diagram from what you know; it enables you in principle to *see* any relationship entailed by what you recalled, however sparse the set of explicitly encoded (i.e., “noticed”) relationships might be. Even if such additional relationships might in principle be inferable from what you recalled, making them explicit without drawing the map might be extremely difficult. One might say that in most cases spatial relations are best represented spatially in a diagram in order to exploit the extraordinary capacity that vision has for extracting spatial relationships. But what about the proposal that something very similar occurs even when the figure is not externalized on paper, but internalized as a mental image or “mental model”? It is to this question that I now turn.

8.4 Thinking with *mental* diagrams

8.4.1 Using mental images

Given the advantages of visual displays in aiding certain kinds of reasoning, it is very tempting to think that mental diagrams would do the same, especially since it feels like you are looking at a diagram when you imagine it. Indeed, many of the scholars who have studied the role of diagrams in reasoning have taken it for granted that their analysis applies to mental diagrams as well, so that visual thinking has come to mean thinking with the aid of visualized (mental) diagrams (Jamnek, 2001). There is no denying that in some ways mental images may help reasoning; the problem is to say in what ways they might help reasoning despite the fact that many of the features of visually augmented reasoning discussed in the previous section do not apply to images. For example, as I argued in the last chapter, you cannot visually perceive new patterns in images (except for simple ones that you arguably “figure out” rather than perceive using the visual system); you do not have a special facility to detect such properties as region inclusion or intersection (as in Figure 8-3 or Figure 8-5), you cannot “see” properties in an image you did not independently know were there, or detect the similarity of figures constructed as a result of drawing certain new lines, as in Figure 8-4; you cannot rely on being able to access several items at once to detect relational properties (such as collinearity and

²⁹ Even if you had a so-called “photographic memory” (eidetic memory) your memory representation would be incomplete in very many ways. Very few people believe that accurate long-term photographic memory exists and even if it did, information-theoretic considerations dictate that it would have to be an approximation in any case. Evidence from studies of visual memory show that it tends to be vague in qualitative ways (see Chapter 1), rather than just fuzzy or low-grade the way a poor TV picture would be approximate. There are a number of studies of Eiditikers and few of the claims of detailed photographic memory have stood up to repeated tests. The one principal study by (Stromeyer & Psotka, 1970) has not been replicated and the search for true eidetic imagery remains elusive (Haber, 1979).

enclosure); and you could not construct a set of distinct cases, such as those in section 8.3.4 and 8.3.5, and use them as an extended memory (since the “internal” memory of your mental image is severely limited).

We saw (in chapter 1, and in various places in the last 3 chapters) that there is a great deal of evidence showing that both visual percepts and imagined pictures are very different from physical drawings. When we imagine carrying out certain operations on mental images (e.g., scanning them, or drawing a line on them) what we “see” happening in our mind’s eye is just what we believe would happen if we were looking at a scene and carried out the real operation, nothing more or less than that. As a result, we cannot rely on discovering something by observing what actually happens, as we might if we were drawing on a piece of paper. Our mental picture also does not have the benefit of being a rigid surface so it does not have the stability and invariance of properties that a physical picture would have when various operations are applied to them. For example unlike a physical diagram, a mental image does not automatically retain its rigid shape when it is transformed, say by rotating it or moving parts of it around or by folding it over, or adding new elements to it. This is because there is no actual inner drawing surface to give an image its rigidity and permanence and because “noticing” new visual patterns in an image is largely an illusion (see Chapter 7).

Just to remind you of the ephemeral quality of image properties, try to imagine a diagram in which you do not know in advance where the elements will fall. For example, imagine an arbitrary triangle and draw lines from each vertex to the midpoint of the opposite side. Do the three lines you draw this way cross at a single common point, or do they form a small triangle inside the original triangle (and is this remain true for any triangle whatsoever)? Imagine a triangle drawn between a pair of parallel horizontal lines (with its base on the bottom parallel and its remaining vertex on the top parallel) and transform the triangle by sliding the top vertex along the line it is on until the triangle becomes highly skewed. (What happens to its area?) Now draw a vertical line from the top vertex to the bottom parallel and two other vertical lines from the bottom vertices of the triangle to the top parallel. How many triangles are there now, and are any of them identical? Imagine a string wrapped around two partially embedded nails, and held taught by a pencil (so the string makes a triangle). Now move the pencil around wherever it can go while keeping the string taught and let it draw a figure on the table: What does this figure look like? Or, to repeat an example I cited earlier; imagine a pair of identical parallelograms, one above the other, with corresponding vertices of the two parallelograms joined by vertical lines. Look at it carefully and note what it looks like. Now draw it and look at the drawing (and do the same with the other examples – draw them and see how they look compared with how they looked in your image).

Jeff Hinton (Hinton, 1987) provides another example of how images lack many of the visual properties that one naturally attributes to them. His example focuses on the concept of horizontal and vertical. Imagine a cube (about one foot on each edge). Place a finger of one hand on one of the vertices and a finger of your other hand on the diagonally opposite vertex. Now hold the cube between these two fingers in such a way that the two vertices are directly above one another (and the cube is now being held with its diagonal vertical). Now without first thinking about it, examine your image and count the remaining vertices of the figure as they appear in your image, and say where they are located relative to the two you are holding (it may help to put the lower vertex down on a table and use your free hand to point to the free vertices). Before reading on you should try this on yourself. Most people get the answer terribly wrong: they end up saying that there are four remaining vertices that lie in a plane, which of course leaves the cube with only 6 vertices whereas we know it has 8. Hinton argues that in rotating the imagined cube we must alter a description of it in order to conclude what it would look like in the new orientation, and in doing so we are deceived by thinking of the solid not as a cube but as a figure that is symmetrical about the new vertical axis.

We saw in chapter 7 that mental figures are not *visually* reinterpreted after we make changes or combine several figures, except in cases where the figures are sufficiently simple that one can figure out (by reasoning) what familiar shape they would look like. If we were to try to prove Pythagoras’ Theorem by constructing the squares in our mind (as in Figure 8-4), we would not be able to use our power of recognition

to notice that there were 4 identical triangles, one in each corner – a recognition that is critical to determining how to proceed with the proof. Some people explain such differences between pictures and mental images by saying that images are limited in their capacity. But while the capacity of mental imagery clearly is limited, the limitation is not one we can understand in terms of how many lines can be drawn, but in terms of how many conceptual units there are (as we saw in the examples of section 1.4.3). Moreover, such limitations are not compatible with the claims often made about the role of more complex images in reasoning, including a large number of experiments involving operations like scanning, rotation and changing the size of images.

The point here is the one that occupied the last 3 chapters of this book: namely, it is a mistake to think that when we are visualizing, there is something in our head that is being seen or is being *visually* interpreted. But if mental diagrams are so different from real diagrams, why then is it easier to do a problem such as the one I discussed earlier, illustrated in Figure 8-5, by imagining that figure in one's "mind's eye"? There is no doubt that it would be very difficult to do this extremely simple problem if one were barred from using a mental image, regardless of what form the image actually takes.³⁰ Why should this be so if the image does not have essentially pictorial properties?

8.4.2 Using mental models

One might reply that it may be a mistake to think of these diagram-like reasoning aids as actual figures, since the only properties that are used are the relative locations of items and their names. Yet there may still be some sort of representation of a spatial layout that is important for reasoning. Such as spatial layout is all that is used in what (Johnson-Laird, 2001) calls "mental models" (see section 8.3.5). Although this avoids some of the problems of viewing the inner diagrams as things that are "seen", it leaves many of the other serious problems intact. In particular in order for the inner mental models to take up some of the functions of external spatial models, we have to assume that mental models have some appropriate properties of stability and (parallel?) accessibility of relational information that are similar to those of external spatial models. Yet, as mentioned in connection with the discussion of what it could mean to say we "think in pictures," much has to be going on off-stage and independent of the way the model appears when drawn on paper. For example when we set out the articles in the example discussed earlier in section 8.3.5 (involving the locations of a spoon, fork, plate and cup) certain assumptions were implicit: that when you place the cup in front of the knife the relative location of the spoon and plate remains unchanged, and that when told that the plate is to the right of the spoon, two possibilities have to be considered, as in layouts labeled (6), in neither of which are the relative locations of the already placed items altered. Well, not quite. Look at the layouts in question, reproduced below (with original numbering) for easier reference.

The description was:

- (1) The spoon is to the left of the knife
- (2) The plate is to the right of the spoon
- (3) The fork is in front of the spoon
- (4) The cup is in front of the knife

And the resulting layout was:

(6)	spoon	knife	plate	OR	spoon	plate	knife
	fork		cup		fork		cup

³⁰ Notice that even here one needs a broader understanding of a "mental image" than the picture theory provides, since blind people can solve such problems without the experience of "seeing". Presumably other sensory modalities, or even unconscious spatial constructions of some sort could also be involved and therefore ought to be accommodated in our understanding of what it means to "imagine".

You will note that one possible way to place the plate to the right of the spoon is to intersperse the plate between the spoon and the knife, as was done in the right-hand part of the pair of figures in (6). The fact that this did not alter the previously encoded relation between spoon and knife, and between the knife and the cup, is attributable to the user's knowledge of both the formal relation "to the right of" (e.g., that it survives interpolation) and the pairwise cohesiveness of the knife-cup pair which were moved together because their relation was previously specified. Properties such as pairwise cohesiveness can be easily seen in a diagram where all that is required in order to place the plate to the right of the spoon is to move the plate-cup column over by one. This movement of a column is easy to understand in the diagram, but because rigidity of form (or constancy of pattern with rigid movement) is not a property of images, one would have separately compute what should happen to parts of a pattern when other parts are changed. This would depend on additional knowledge of when some things can change while other things remain unchanged (e.g., so-called frame axioms of computational reasoning). Consequently, in order for the mental version of the model to give the results that would be observed in the physical layout case, additional reasoning would be required, as well as additional memory requirements. "Mental models" like mental images, are assumed to have certain unstated properties because if they were drawn on paper, as they are when they are being explained, the properties would hold automatically by virtue of physical properties of the surface and properties of the visual system.

8.4.3 What happens during visualization?

If images (or mental models) have such serious inherent problems, why then do they appear to help – at least in some circumstances – in reasoning? Perhaps the reason does not lie in any of the intrinsic properties of images or models that are tacitly assumed. Perhaps it is because of the content of the image or model, and not its alleged spatial form, or the architecture of the imagery or visual system. It is a general property of mental life that we tend to think different thoughts (thoughts with different contents) when we are presented with a problem in different ways (as Hayes & Simon, 1976, showed clearly). In fact a major step in problem-solving is formulating or representing the problem in an appropriate way (Herb Simon goes so far as to claim that "Solving a problem simply means representing it so as to make the solution transparent," Simon, 1969). It appears that we think about different things (or at least we highlight different properties) when we *visualize* a situation than when we do not, and so perhaps the thoughts that arise under these conditions are helpful in solving certain kinds of problems, especially problems that concern spatial patterns. So part of the answer for why imagining a figure is sometimes helpful may simply be that when we think in terms of spatial layouts we tend to think about different things and represent the problem in different ways from when we think in terms of abstract properties. Such thinking is rightly thought of as being more concrete because its contents (as opposed to its form) concern spatio-temporal and visual properties rather than abstract properties. When we think about a rectangle in the abstract we cannot readily understand the instruction to draw lines from the two bottom vertices to a point on the opposite side, because we may not have individuated these vertices: you can think "rectangle" without also thinking of each of the 4 vertices, but when you visualize a rectangle you do think of each of its 4 vertices.

My experience in using the example of Figure 8-5 in class is that the hardest part of solving the problem without using the blackboard, is communicating the problem specifications. The way I am invariably forced to do it is to ask students to imagine a rectangle whose vertices are labeled, starting from the bottom left corner and going clockwise, as A, B, C, and D and then drawing a line from A to any point on CD and from D to any point on AB. By conveying the problem in this way I ensured that the individual vertices will be distinct parts of their representation of the figure, and I also provided a way to refer to these parts individually. This makes the task easier precisely because this way of describing the problem sets up thoughts about a particular rectangle and about particular individuated and labeled vertices and sides. But even then one usually cannot select the side labeled CD without going through the vertex labels in order. Thus it may be that imaging certain figures is helpful in solving problems because this way of thinking about the problem (by

imagining that one is seeing the figure) focuses attention on individual elements and their relational properties. Looking at the role of imagined figures in this way is neutral on the question of the form of the representation of mental images, and indeed may help to reduce the temptation to reify an inner diagram with its implication that the visual system's involvement in the visualization process is that of observing the mental image.

But why should imagery lead to entertaining certain kinds of representational contents, such as those alluded to above? In Chapter 7 (section 7.6.2) I discussed ways in which representations underlying mental images might differ from other sorts of cognitive representations. The main points were that images are a more restricted form of representation because they are confined to representing individual instances rather than the more general situations that can be described using quantified statements: they do not represent the belief that all crows are black but only that crow #1 is black, crow #2 is black, crow #3 is black, and so on. They individuate elements of a situation and may represent parts of a scene by distinct parts of the representation. They also tend to attribute properties to particular elements (as in, crow #1 is black). They do not contain quantifiers or negation, except by the addition of annotations that have to be interpreted by mechanisms outside the imagery system (by the sorts of mechanisms that interpret symbolic expressions). Thus such representations are more like a relational database or a restricted representational language, like Levesque's "vivid representation" (Levesque, 1986). Beyond such properties, which images do share with pictures (but also with relational databases), there are few pictorial properties that withstand careful scrutiny, including the widely believed spatial nature of images. While individuating elements could involve representing them in distinct physical locations in the brain, the locations and relations among these locations are no more signifying than are locations of variables in a computer, or of the physical location of cells in a matrix data structure.

8.5 Imagery and imagination

One of the most interesting questions one might ask about mental imagery is how it serves as a vehicle for creative imagination, for there is no doubt that creativity often starts with imagining things that do not exist (or, in the case of some "thought experiments" *cannot* exist). Before considering this question, I want to point out a systematic ambiguity in discussions about imagination. There are two senses in which we can talk about imagining something. One is the sense in which a person might imagine *seeing* something (often referred to as "imaging" or as "visualizing"). The other is the sense in which a person simply considers a (counterfactual) situation in which something might be true. This is the difference between "imagining X" and "imagining *that X is the case*", or between imagining seeing X and considering the implications of X being the case (as in "what if X were true?"). Clearly, the latter is essential not only for creative invention, but also for mundane planning. In order to make a plan one needs to consider non-actual (or counterfactual or fictitious) situations; one must be able to think "If I did this then it would result in such-and-such," which entails considering (and therefore representing) some non-actual state of affairs. There is much that can be said about this subjunctive or imagining-that form of reasoning (see, e.g., Harper, Stalnaker, & Pearce, 1981) but it clearly does not require generating a mental image; it does not require "seeing in the mind's eye". The other sense of imagining does require entertaining such a mental image since it in effect means "imagine what it would look like if you were to see...." some particular situation. What exactly it means, in terms of information processing mechanisms and operations, to imagine that you were seeing something is what this discussion of mental imagery has been about. Whatever else it means, we know is that it does not mean that we have a picture of the situation displayed in our brain. We also have good reason to think that imaging in that sense involves at least some form of reasoning (using our tacit knowledge of principles as well as our memory or relevant past events) concerning what would happen in the imagined situation and that it also involves entertaining some symbolic representation of the relevant counterfactual situation (with or without an analogue representation of magnitudes). Beyond that there is not much we can say about this process that would cast light on its role in creative visualization.

Libraries are full of books about the creativity of visual thinking. The history of science abounds in examples of important discoveries being made when a scientist thinks about a problem by visualizing (this idea is discussed very widely – see Shepard, 1978a, for a good summary of this point with extensive biographical citations). Scientists as illustrious as Albert Einstein reported that he thought in images. How do these reports square with what I have been saying about the non-pictorial nature of imagery and especially of the non-pictorial nature of thinking and reasoning? Are these reports simply misled or romanticized? The simple answer is that what people are reporting does not constitute a theory of what goes on in the mind that is causally responsible for episodes of creative thinking. It is thus neutral on the question of whether thinking using mental images consists of examining pictorial entities in one's head. As I have said many times throughout this book, such reports describe what people are thinking *about*, not what they are thinking *with* and in any case they are not reports of causal processes underlying the experiences. For example, in the very widely cited case of Albert Einstein, what Einstein claimed was that (a) his thoughts were not expressed in words, at least not initially, (b) his discoveries were not made in a strictly logical manner, and (c) many of his ideas were developed in the course of carrying out “thought experiments” in which, for example, he imagined himself traveling along the front of a beam of light at the speed of light. Interpreting this to mean that a “visual” mode of reasoning is more creative than other modes requires acceptance of a naïve version of the dual code idea (i.e., what is not logical is not verbal and what is not verbal is pictorial). It also requires conflating two senses of “imagine” that I mentioned at the beginning of this section; a sense that means “consider what it would be like if x were true,” and a sense that means “imagine seeing x happen”. Evidence that scientists do not think “in words” or that they do not discover new theories by following (deductive) logic tells us nothing about the process that they do go through – nobody thinks in “words,” nor do they make explicit use only of deductive logic in forming hypothesis. To say that they do it “by intuition” is not to say that it involves pictorial mental images, it is just another way of saying that we have no idea how it happens.

While admitting our current state of ignorance about such things as where creative ideas come from, it is worthwhile to keep in mind where the real mystery lies. It lies not in questions about the nature or format of mental images, but in questions about the nature of creative thought, and in particular in the question of where new ideas come from. Whatever their underlying form, both imagery and language are at most a vehicle for expressing ideas. They cannot explain the origins of the ideas, much as it may be tempting to think that we simply “notice” something new or are “inspired” to have novel thoughts by observing our images unfold. The fact that a new idea may arrive accompanied by a particular image (e.g., Kekulé's dream of the snake swallowing its tail, which suggested the shape of the benzene molecule) is not the central fact; what is central is the idea which caused this image in the first place and we have not the slightest notion where that came from. Recalling, or calling to mind, is equally mysterious whether or not it is accompanied by the experience of a mental image or of inner dialogue (especially since neither form is capable of fully expressing our thoughts). The source of our ideas is, as Kant put it (Kant, 1998), “... an art, hidden in the depths of the human soul, whose true modes of action we shall only with difficulty discover and unveil.” But that is no reason to substitute one mystery (the role of imagery in creative insights) for another (where new thoughts come from).

Most writings about the role of imagery in creative problem solving, as well as in psychotherapy, typically do not make assumptions about the medium or the format of images, nor need they even take a position on whether thoughts experienced as images are different in any way from thoughts experienced as dialogue, *except for what they are about*. Clearly when we imagine or visualize a scene we are attending to different properties than we would if we were to, say, formulate a written description of it. The question of whether when we have the experience of “seeing in our mind's eye” we are thinking about different things, whether the subject or the meaning of our thoughts is different, is not in contention. It is obvious that when I think about a painting I am thinking of different things than when I think about mathematics. The difference between the contents in the two cases, though they may seem small when the thoughts are on similar topics,

may be sufficient to explain the why different behaviors result from imagery-thoughts than from non-imagery-thoughts. For example I can think about my office room with or without visualizing it. But when I visualize it, the subject matter of my thought includes the appearance of the room and the appearance of some of the individual things in it. When I think about my office room without visualizing it I have selected certain aspects of the room to be the subject of my thoughts. While I can think some thoughts about specific visual features, such as where the chair is located, without visualizing the room, the subject matter is still different. The thought may not contain explicit references to colors or sizes or shapes. But this difference is still a content difference and not a form difference. The same is true of other uses of mental imagery; say in problem solving, as we saw in the previous section. The difference between thinking about geometrical theorems with or without visualizing is largely that in the latter case I am actually thinking of the details of a *particular token* figure (a particular triangle or diagram). The interesting question of whether this entails a different form of representation is not addressed by the fact that in one case I visualize something and in another I do not.

One of the reasons why imagery is widely thought to contain the secret of creative imagination lies in its apparent autonomy and freedom from conscious ratiocination; it often seems to us that we “see” new meanings in images or that new ideas are “suggested” to us in a quasi visual mode, and that this process follows a different logic, or perhaps no logic at all but some other intuitive principles.³¹ The terminology of “seeing” is frequently used to indicate a new way of conceptualizing a problem or idea (as in “Now I see what you mean”). It has even been suggested that the process of *seeing as* hides the secret of the role of imagery in creative imagination. But as I argued earlier (and in Pylyshyn, 1999) *seeing as* is the clearest example of an aspect of visual perception that is not constitutive of and unique to vision. Unlike the processes of early vision, *seeing as* occurs in the stage of visual perception that is shared with cognition generally. *Seeing as* involves not only seeing, but also reasoning, recalling and recognizing. It is the part of the visual process where belief fixation occurs, as opposed to the part of the process in which the appearance of a scene as a set of three-dimensional surfaces is computed, and is consequently where vision crosses over to memory, recognition, inference, decision making and problem solving. Logical and quasi-logical reasoning occurs with every mental act we perform, including seeing something as a member of some particular category. It is not the case, as many writers have assumed, that the sorts of semantically-coherent processes that operate over beliefs and that lead to new beliefs, are capable of only rigid logical reasoning (for a critique of the use of logic to model intelligence, see McDermott, 1989). While processes that I refer to as “reasoning” do respect the meaning of representations and therefore lead to thoughts whose content has some connection to the content of other thoughts that trigger them, they need not be confined to logical or rational processes; semantic coherence is also true of creative innovation. The difference is that while we know something about rational reasoning (thanks to work on the foundations of logic over the past century) we know next to nothing about non-demonstrative, abductive, or common-sense reasoning and even less about creative reasoning. However none of this ignorance is aided by talk of the role of the mind’s eye in creativity since that just substitutes one mystery for another.

Having said this, I will nonetheless add my speculative ideas to the many others around in considering ways in which the imagery system might take part in creative reasoning and problem solving, for it is clear that there is something special about what we can do when we are imagining that is different from what we can do when we are not.

³¹ The notion that ideas interact according to special mental principles is close to the British Empiricist notion of “mental chemistry,” according to which the formation of concepts and the flow of thought is governed by principles such as temporal and spatial contiguity, repetition, similarity, and vividness, which were believed to favor the formation of associations.

8.5.1 How does creativity connect with imagery?

One idea that plays a role in virtually every theory of creative thinking is the notion that at some stage thinking makes a non-logical leap; reasoning happens, as has sometimes been said, “outside the box”. Of course not anything outside the box is creative reasoning, some of it is just crazy. The creative step must not merely be a non sequitur, but it must be related somehow to the topic at hand. So anything that can encourage nonlogical, yet content-related progression of thoughts, could conceivably serve as a stage or a catalyst for creative thinking. Indeed several theories explicitly posit stages in which representations are altered in random ways and many methods that teach creative thinking using imagery do encourage people to come up with nonlogical connections, to create associations based on similarities and to follow leads based on ideas suggested by the images. Random perturbations by themselves serve no function (as they say in the computer field, *Garbage In, Garbage Out*). But randomness in conjunction with selectivity can, at least in principal, lead to some creative leaps. Yet this too is not enough. The proverbial thousand monkeys at typewriters will never produce Shakespearian prose, even if the probability is actually finite. What is needed for a genuine creative product (other than the skill to produce it) is some content-restricted thought transitions that are non-deductive, in the sense that they are not like valid proofs, but are more like abductive conjectures. People have looked to imagery as a mechanism that might allow transitions that are non-logical (or perhaps *alogical*). For example, it has been suggested that when we examine a drawing or sketch that we have made based on ideas that are relevant to the problem at hand, we can “notice” consequences that we might not be able to infer explicitly, or we might be reminded of something by virtue of its similarity to the object depicted in the drawing, or we might use the drawing as a metaphor for some abstract concept with which we are struggling, and so on. This kind of visual aid provided by one’s own sketches is unproblematic, as we saw in section 8.3.6, and it happens very often (some people carry around a sketchpad for just this purpose). Inferences carried out by noticing properties in a diagram that were implicit in the knowledge from which the diagram was constructed can be understood on the same basis as those that explain why diagrams are useful aids in thinking (see the discussion of this point at the beginning of this chapter, e.g., section 8.3).

Despite the caveats I have offered concerning the significant differences between drawings and mental images, it could still be the case that the mere fact that one is visualizing (whatever that actually means in terms of information processes) brings to the fore certain properties of the object of our thoughts that are related to shape and appearance. After all, whatever else happens when we visualize, it is a matter of definition that we consider the appearance of the objects we visualize. So even though there need be no difference in the form of the information in thought and vision, the difference in content and on the salience of difference contents, when we visualize could mean that different principles come into play in determining what we think of next. For example, thinking about the shape of particular figures may remind you of other figures that have a similar shape, so it may be that the “reminders” run along a shape-similarity dimension rather than some other dimension that would be more salient when the subject of the thoughts is something other than appearances. This emphasis on shape and shape-related relationships may even carry with it some of the same characteristics as those we observe when we perceive actual diagrams. So, for example, one can think about sets of things that have certain properties and one can think about how having certain properties entails having other properties, yet the reasoning process may be much easier if we think, instead, of the properties of intersection of circles in Venn diagrams *even if the mental representations do not themselves have circles, areas, overlapping regions* and so on. Just thinking about these properties in whatever form we think of them (e.g., in a symbolic language of thought) can still give us some of the benefit we would have if we had actual visible Venn diagrams. This can be true as long as it is easier (for whatever reason) to think about relations among regions and intersections between regions, than of other properties of the same formal type.

Roger Shepard (Shepard, 1978a) has given a great deal of thought to the question of why images should be effective in creative thought and invention. He lists four general properties of imagery that make them

especially suitable for creative thought. These are (pp 156): “their private and therefore not socially, conventionally, or institutionally controlled nature; their richly concrete and isomorphic structure; their engagement of highly developed, innate mechanisms of spatial intuition; and their direct emotional impact.” These are perfectly reasonable speculations, with varying plausibility. The first general property (imagery is more private) rests on the dual code assumption since there is no reason why a pictorial form of representation is any more immune from external control than the form that I have discussed in this book (i.e., a conceptual system or “language of thought”). The second property (isomorphic structure) relies on a theory of imagery that suffers from being insufficiently constraining. As we saw in section 6.3.2.3, the sort of isomorphism that Shepard has in mind (what he calls “second order isomorphism”) is a requirement for any adequate representation and leaves open the crucial question of whether the isomorphism arises from intrinsic properties of the representation or from inferences based on tacit knowledge. Consequently, while the claim that images have a concrete and isomorphic structure is a reasonable one, it does not favor one theory of image representation over another, although it does suggest that perhaps thinking in terms of concrete (token) instances may be an effective way to induce creative leaps. The third property (images engage mechanisms of spatial intuition) is also reasonable, and in fact is in accord with proposals I made earlier in section 7.3 where I discuss reasons why mental images appear to have “spatial” properties. There is no doubt that our spatial sense is extremely central in how we cognize the world and may be particularly relevant when we think about spatial patterns, shapes, and the operation of physical mechanisms. But once again, the impression that we “see” things is misleading. What we do, according to the analysis presented in Chapter 7, is think of individual things as located in a space that we are currently perceiving, either visually or proprioceptively. As I argued in section 7.3.2, we may use the real physical space that we perceive with one or another sense as a receptacle into which we can project spatial patterns that we are thinking about, and in this way make use of the “highly developed, innate mechanisms of spatial intuition” that Shepard speaks of. In fact, we can even use this spatial sense in order to individuate and keep track of abstract ideas, as is done in American Sign Language.

The last property that Shepard proposed as a basis for the superiority of mental images over words in promoting creative thought is their “direct emotional impact.” Once again this is a reasonable proposal that does not favor one theory of the nature of mental images over another. It is true that if we think the sentence “I am about to be hit by a falling brick” we are not nearly as moved to action, or to emotion, as we are if we were to visualize a brick is falling down on us. While there has not been a great deal of emphasis placed on this aspect of mental imagery, it is surely very important. The power of images to create states of fear, anxiety, arousal, or sudden flinching or recoil (as when one imagines oneself at the edge of a precipice), is a property that in the past (e.g., as I was writing my first critique of mental imagery, Pylyshyn, 1973) had been made it most difficult for me to fully accept the symbolic view of mental imagery. Dealing with this power of imagery requires continually reminding oneself that the crucial parallel between seeing and emotional states, on the one hand, and imagery and emotional states on the other, very likely arises from common properties of the two types of mental representations, neither of which has anything to do with their having a pictorial format. While they clearly both possess a vividness that may be absent in more deliberate forms of ratiocination, this vividness does not argue for one format of representation over another. The vividness may have more to do with the fact that images and visual representations both focus on the particular – on particular objects and individual episodes. Mental representations are powerful forces in our lives and representations of *particular* things and events are more powerful to us as individuals than are representations of general principles.

8.5.2 Enhancing creative thinking

Creativity is much prized in our culture. It is also much misunderstood (for example it is associated with being a misfit, with nearness to insanity and so on). Measures of intelligence and of creativity, despite their many shortcomings, have generally shown the two to be highly related. This is not surprising since it is hard to be creative if you don’t understand the problem or are unable to reason about it well. Because creativity is

especially prized in certain fields (e.g., science and art), people have tried to find specific factors that are associated with high creativity, including personality traits, styles of problem-solving, and strategies that enhance creative expression. Such factors are bound to be there since we do have at least some vague idea of what distinguishes creativity from mere ability. Some of the studies have led to interesting findings and have connected creativity to nonstandard ways of thinking or nonstandard values (e.g., a playful attitude towards problems and a high value placed on finding and formulating problems for its own sake, see Getzels & Jackson, 1962). In any case what we consider to be creative solutions to a problem often require that the problem-solver generate a space of options that go beyond the expected or the obvious. For this reason many creativity-enhancing techniques encourage free association, exploration of unusual similarities and relationships among elements of the problem, and other tricks to help bring to the fore aspects that might not readily come to mind in connection with the problem at hand – all in the interest of temporarily subverting habitual modes of thought and reasoning.

An example of the use of mental imagery to generate associations and to trigger alternative ideas in this way is the game that (Finke, 1990) developed to study the enhancement of creativity through mental imagery. This game asks people to make a sketch of what they think of when they think about a certain problem (or just to sketch some object involved in that problem) and then to modify it in certain ways or just think of what it reminds you of. The game relies in the fact that the manipulation of shapes and the recognition of similarity of shape can remind one of things that are not connected to the problem in a logical goal-directed way, and thereby may influence the sequence of thoughts one goes through in coming up with an idea. There are many examples of the use of such techniques which rely on mental imagery to enhance creativity – in fact the World Wide Web is filled with “creativity resources” consisting of everything from computer programs to actual physical devices, all designed to inhibit one’s normal problem-solving habits.

Much of the advice for enhancing creative problem solving is focused quite specifically on encouraging nonstandard ways of thinking about a problem by emphasizing the association of ideas over more direct ways of working on a problem, such as deduction and means-ends analysis. In fact some of the tools that have been claimed to enhance creativity are just tools for inhibiting the usual habits of problem solving from controlling the process. (Gardner, 1982, p27) describes what he calls “whimsical” devices for breaking away from “vertical thinking” into “lateral thinking”. For example, there is a device called the “Think Tank” which is a rotating drum containing 13,000 tiny plastic chips, each with a word printed on it. When the drum is rotated, different words come into view and the user is invited to free associate to each word or phrase. Related tools are being sold today that run on computers and produce graphic pictures that allow you to connect ideas (e.g., the “Mind Mapper”). Some years ago there was even a special room, called the Imaginarium (McKim, 1978), for training people to think creatively by leading them through a sequence from relaxation, through imagining the superposition and transformation of images (done with the help of slides and sounds), through a series of fantasies involving metaphors, to Zen-like attention to the here-and-now, coupled with physical exercises. What all these ideas and methods have in common is this: They tend to encourage you to do things in a different way from how you normally would have done them (and perhaps even to feel different about what you are doing).

Of course not any different way is a good way, and inhibiting the usual ways of solving a problem need not lead to an improvement. But when you are stuck, it may well be a useful starting heuristic.³² And so it is with many uses of imagery. Free-associating to images may be different from associating to words, but association is no way to enhance creativity. Or is it? After looking at a wide range of claims about how imagery enhances creativity it seems to me to be a mixed bag: it includes everything from advice on the use of graphics and other tools to the equivalent of giving a thousand monkeys graphics software instead of typewriters. It is safe to say that in the present state of our understanding of both creativity and mental imagery, this heterogeneous collection of talismans tells us little about the nature of creativity, and even if imagery techniques did enhance one's creativity it would be no more informative than the discovery of the power of placebos to alter one's somatic complaints. They may work, but they tell us nothing about the underlying mechanisms.

REFERENCES

- Allwein, G., & Barwise, J. (Eds.). (1996). *Logical reasoning with diagrams*. New York: Oxford University Press.
- Anderson, B., & Johnson, W. (1966). Two Kinds of Set in Problem Solving. *Psychological Reports*, 19(3), 851-858.
- Anderson, J. R. (1978). Argument Concerning Representations for Mental Imagery. *Psychological Review*, 85, 249-277.
- Anderson, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, 29, 303-330.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61, 183-193.
- Attneave, F., & Farrar, P. (1977). The visual world behind the head. *American Journal of Psychology*, 90(4), 549-563.
- Attneave, F., & Pierce, C. R. (1978). Accuracy of extrapolating a pointer into perceived and imagined space. *American Journal of Psychology*, 91(3), 371-387.
- Avant, L. L. (1965). Vision in the ganzfeld. *Psychological Bulletin*, 64, 246-258.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723-767.
- Banks, W. P. (1981). Assessing relations between imagery and perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 844-847.
- Barolo, E., Masini, R., & Antonietti, A. (1990). Mental rotation of solid objects and problem-solving in sighted and blind subjects. *Journal of Mental Imagery*, 14(3-4), 65-74.
- Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577-660.
- Bartolomeo, P. (2002). The relationship between visual perception and visual mental imagery: A reappraisal of the neuropsychological evidence. *Cortex*, 38(3), 357-378.
- Bartolomeo, P., Bachoud-levi, A. C., & Denes, G. (1997). Preserved imagery for colours in a patient with cerebral achromatopsia. *Cortex*, 33(2), 369-378.
- Bartolomeo, P., & Chokron, S. (2002). Orienting of attention in left unilateral neglect. *Neuroscience and Biobehavioral Reviews*, 26(2), 217-234.

³² There is a technique used in servomechanisms, which may also been seen in some thermostats, that allows for more precise control by inhibiting static friction, which is always higher than dynamic friction (it takes more force to get something static to move than to change its velocity once it has started to move). The technique is to introduce some random perturbation into the control loop, using what is sometimes called a "dithering device". In a thermostat one can sometimes here a buzz that keeps the contact points from sticking and in military applications involving the control of heavy arms, it can be introduced as noise in the position controller. Search processes that seek to maximize some quantity using a method such as "steepest ascent", can become locked in a local maximum and thereby miss the optimum global maximum. It has been suggested that problem solving and thinking could use a dithering device to keep people from getting stuck in a "local maximum." Perhaps the sorts of creativity-enhancing ideas discussed above are just varieties of dithering devices.

- Barwise, J., & Etchemendy, J. (1990). Information, inferences, and inference. In R. Cooper, K. Mukai & J. Perry (Eds.), *Situation theory and its applications, I*. Chicago, IL: University of Chicago Press.
- Behrmann, M. (2000). The mind's eye mapped onto the brain's matter. *Current Directions in Psychological Science*, 9(2), 50 - 54.
- Behrmann, M., Moscovitch, M., & Winocur, G. (1994). Intact visual imagery and impaired visual perception in a patient with visual agnosia. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1068-1087.
- Behrmann, M., & Tipper, S. (1999a). Attention accesses multiple reference frames: Evidence from unilateral neglect. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 83-101.
- Behrmann, M., & Tipper, S. P. (1999b). Attention accesses multiple reference frames: Evidence from visual neglect. *Journal of Experimental Psychology: Human Perception & Performance*, 25(1), 83-101.
- Behrmann, M., Winocur, G., & Moscovitch, M. (1992). Dissociation between mental imagery and object recognition in a brain-damaged patient. *Nature*, 359(6396), 636-637.
- Berger, G. H., & Gaunitz, S. C. (1977). Self-rated imagery and vividness of task pictures in relation to visual memory. *British Journal of Psychology*, 68(3), 283-288.
- Bernbaum, K., & Chung, C. S. (1981). Müller-Lyer illusion induced by imagination. *Journal of Mental Imagery*, 5(1), 125-128.
- Beschin, N., Basso, A., & Sala, S. D. (2000). Perceiving left and imagining right: Dissociation in neglect. *Cortex*, 36(3), 401-414.
- Beschin, N., Cocchini, G., Della Sala, S., & Logie, R. H. (1997). What the eyes perceive, the brain ignores: a case of pure unilateral representational neglect. *Cortex*, 33(1), 3-26.
- Biederman, I. (1987). Recognition-by-components: A theory of human image interpretation. *Psychological Review*, 94, 115-148.
- Biederman, I. (1995). Visual object recognition. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual Cognition* (second ed.). Cambridge, MA: MIT Press.
- Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research*, 39(17), 2885-2899.
- Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *JEP:HPP*, 18, 121-133.
- Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance*, 19(6), 1162-1182.
- Binford, T. O. (1971, December). *Visual perception by computer*. Paper presented at the IEEE Systems Science and Cybernetics Conference, Miami.
- Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, 14(1), 129-133.
- Bisiach, E., Ricci, R., Lualdi, M., & Colombo, M. R. (1998). Perceptual and response bias in unilateral neglect: Two modified versions of the Milner Landmark task. *Brain & Cognition*, 37(3), 369-386.
- Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is the richness of the visual world an illusion? Transsaccadic memory for complex scenes. *Perception*, 24(9), 1075-1081.
- Block, N. (1981a). Introduction: What is the issue? In N. Block (Ed.), *Imagery* (pp. 1-16). Cambridge, MA: MIT Press.
- Block, N. (2001). Is the content of experience the same as the content of thought? In D. Emmanuel (Ed.), *Language, Brain, and Cognitive Development: Essays in honor of Jacques Mehler*. Cambridge, MA: MIT Press/Bradford Books.
- Block, N. (Ed.). (1981b). *Imagery*. Cambridge, Mass.: MIT Press, a Bradford Book.
- Bolles, R. C. (1969). The role of eye movements in the Muller-Lyer illusion. *Perception & Psychophysics*, 6(3), 175-176.
- Bower, G. H., & Glass, A. L. (1976). Structural units and the reintegrative power of picture fragments. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 456-466.
- Brigell, M., Uhlarik, J., & Goldhorn, P. (1977). Contextual influence on judgments of linear extent. *Journal of Experimental Psychology: Human Perception & Performance*, 3(1), 105-118.
- Broerse, J., & Crassini, B. (1981). Misinterpretations of imagery-induced McCollough effects: A reply to Finke. *Perception and Psychophysics*, 30, 96-98.

- Broerse, J., & Crassini, B. (1984). Investigations of perception and imagery using CAEs: The role of experimental design and psychophysical method. *Perception & Psychophysics*, 35(2), 155-164.
- Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 22(5), 349-368.
- Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, 84(1), 141-147.
- Canon, L. K. (1971). Directed attention and maladaptive "adaptation" to displacement of the visual field. *Journal of Experimental Psychology*, 88(3), 403-408.
- Carpenter, P. A., & Eisenberg, P. (1978). Mental rotation and the frame of reference in blind and sighted individuals. *Perception & Psychophysics*, 23(2), 117-124.
- Casey, E. (1976). *Imagining: A phenomenological Study*. Bloomington, IN: Indiana University Press.
- Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental psychology*, 11, 317-328.
- Chara, P. J., & Hamm, D. A. (1989). An inquiry into the construct validity of the Vividness of Visual Imagery Questionnaire. *Perceptual & Motor Skills*, 69(1), 127-136.
- Charlot, V., Tzourio, N., Zilbovicius, M., Mazoyer, B., & Denis, M. (1992). Different mental imagery abilities result in different regional cerebral blood flow activation patterns during cognitive tasks. *Neuropsychologia*, 30(6), 565-580.
- Chatterjee, A., & Southwood, M. H. (1995). Cortical blindness and visual imagery. *Neurology*, 45(12), 2189-2195.
- Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review*, 76(4), 387-404.
- Cocude, M., Mellet, E., & Denis, M. (1999). Visual and mental exploration of visuo-spatial configurations: behavioral and neuroimaging approaches. *Psychol Res*, 62(2-3), 93-106.
- Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., et al. (1996). Changes in cortical activity during mental rotation: A mapping study using functional MRI. *Brain*, 119.
- Coren, S. (1986). An efferent component in the visual perception of direction and extent. *Psychological Review*, 93(4), 391-410.
- Coren, S., & Porac, C. (1983). The creation and reversal of the Mueller-Lyer illusion through attentional manipulation. *Perception*, 12(1), 49-54.
- Cornoldi, C., Bertuccelli, B., Rocchi, P., & Sbrana, B. (1993). Processing capacity limitations in pictorial and spatial representations in the totally congenitally blind. *Cortex*, 29(4), 675-689.
- Cornoldi, C., Calore, D., & Pra-Baldi, A. (1979). Imagery ratings and recall in congenitally blind subjects. *Perceptual & Motor Skills*, 48(2), 627-639.
- Coslett, H. B. (1997). Neglect in vision and visual imagery: a double dissociation. *Brain*, 120, 1163-1171.
- Craig, E. M. (1973). Role of mental imagery in free recall of deaf, blind, and normal subjects. *Journal of Experimental Psychology*, 97(2), 249-253.
- Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, 375(11), 121-123.
- Currie, G. (1995). Visual imagery as the simulation of vision. *Mind & Language*, 10(1-2), 25-44.
- Dalla Barba, G., Rosenthal, V., & Visetti, Y.-V. (2002). The nature of mental imagery: How "null" is the null hypothesis? *Behavioral and Brain Sciences*, 25(2), xxx-xxx.
- Dalman, J. E., Verhagen, W. I. M., & Huygen, P. L. M. (1997). Cortical blindness. *Clinical Neurology & Neurosurgery*(Dec), 282-286.
- Dauterman, W. L. (1973). A study of imagery in the sighted and the blind. *American Foundation for the Blind, Research Bulletin*, 95-167.
- Davies, T. N., & Spencer, J. (1977). An explanation for the Mueller-Lyer illusion. *Perceptual & Motor Skills*, 45(1), 219-224.
- De Soto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality & Social Psychology*, 2(4), 513-521.
- De Vreese, L. P. (1991). Two systems for colour-naming defects: verbal disconnection vs colour imagery disorder. *Neuropsychologia*, 29(1), 1-18.
- Dehaene, S. (1995). Electrophysiological evidence for category-specific word processing in the normal human brain. *Neuroreport: an International Journal for the Rapid Communication of Research in Neuroscience*, 6(16), 2153-2157.

- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79(1-2), 1-37.
- DeLucia, P. R., & Liddell, G. W. (1998). Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. *Journal of Experimental Psychology: Human Perception & Performance*, 24(3), 901-914.
- Démonet, J.-F., Wise, R., & Frackowiak, R. S. J. (1993). Language functions explored in normal subjects by positron emission tomography: A critical review. *Human Brain Mapping*, 1, 39-47.
- Denis, M., & Kosslyn, S. M. (1999). Scanning visual mental images: A window on the mind. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition*, 18(4), 409-465.
- Dennett, D. C. (1978). *Brainstorms*. Cambridge, Mass.: MIT Press, a Bradford Book.
- Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown & Company.
- D'Esposito, M., Detre, J. A., Aguirre, G. K., Stallcup, M., Alsop, D. C., Tippet, L. J., et al. (1997). A functional MRI study of mental image generation. *Neuropsychologia*, 35(5), 725-730.
- Dodds, A. G. (1983). Mental rotation and visual imagery. *Journal of Visual Impairment & Blindness*, 77(1), 16-18.
- Driver, J., & Vuilleumier, P. (2001). Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition*, 79(1-2), 39-88.
- Dunham, W. (1994). *The mathematical universe*. New York: John Wiley & Sons.
- Easton, R. D., & Bentzen, B. L. (1987). Memory for verbally presented routes: A comparison of strategies used by blind and sighted people. *Journal of Visual Impairment & Blindness*, 81(3), 100-105.
- Escher, M. C. (1960). *The Graphic Work of M.C. Escher*. New York: Hawthorn Books.
- Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review*, 95(3), 307-317.
- Farah, M. J. (1989). Mechanisms of imagery-perception interaction. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 203-211.
- Farah, M. J. (1994). Beyond "pet" methodologies to converging evidence. *Trends in Neurosciences*, 17(12), 514-515.
- Farah, M. J. (1995). The neural bases of mental imagery. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 963-975). Cambridge, MA: MIT Press.
- Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, 20(4), 439-462.
- Farah, M. J., Soso, M. J., & Dasheiff, R. M. (1992). Visual angle of the mind's eye before and after unilateral occipital lobectomy. *J Exp Psychol Hum Percept Perform*, 18(1), 241-246.
- Farrell, M. J., & Robertson, I. H. (1998). Mental rotation and the automatic updating of body-centered spatial relationships. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 483-500.
- Festinger, L., White, C. W., & Allyn, M. R. (1968). Eye Movements and Decrement in the Muller-Lyer Illusion. *Perception & Psychophysics*, 3(5-B), 376-382.
- Fidelman, U. (1994). A misleading implication of the metabolism scans of the brain. *Int J Neurosci*, 74(1-4), 105-108.
- Finke, R. A. (1979). The Functional Equivalence of Mental Images and Errors of Movement. *Cognitive Psychology*, 11, 235-264.
- Finke, R. A. (1980). Levels of Equivalence in Imagery and Perception. *Psychological Review*, 87, 113-132.
- Finke, R. A. (1989). *Principles of Mental Imagery*. Cambridge, MA: MIT Press.
- Finke, R. A. (1990). *Creative imagery: Discoveries and inventions in visualization*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Finke, R. A., & Freyd, J. J. (1989). Mental extrapolation and cognitive penetrability: Reply to Ranney and proposals for evaluative criteria. *Journal of Experimental Psychology: General*, 118(4), 403-408.
- Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *J Exp Psychol Hum Percept Perform*, 6(1), 126-139.
- Finke, R. A., & Kurtzman, H. S. (1981a). Mapping the visual field in mental imagery. *J Exp Psychol Gen*, 110(4), 501-517.
- Finke, R. A., & Kurtzman, H. S. (1981b). Methodological considerations in experiments on imagery acuity. *Journal of Experimental Psychology: Human Perception and Performance*, 7(4), 848-855.
- Finke, R. A., & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(2), 142-147.

- Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science*, 13(1), 51-78.
- Finke, R. A., & Schmidt, M. J. (1977). Orientation-specific color aftereffects following imagination. *Journal of Experimental Psychology: Human Perception & Performance*, 3(4), 599-606.
- Fisher, G. H. (1976). Measuring ambiguity. *American Journal of Psychology*, 80, 541-557.
- Fletcher, P. C., Shallice, T., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Brain activity during memory retrieval: The influence of imagery and semantic cueing. *Brain*, 119(5), 1587-1596.
- Fodor, J. A. (1968). The Appeal to Tacit Knowledge in Psychological Explanation. *Journal of Philosophy*, 65, 627-640.
- Fodor, J. A. (1975). *The Language of Thought*. New York: Crowell.
- Fodor, J. A. (1981). Imagistic representation. In N. Block (Ed.), *Imagery*. Cambridge, MA: MIT Press.
- Fodor, J. A. (2001). Language, Thought and Compositionality. *Mind and Language*, 16(1), 1-15.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.
- Ford, K., & Pylyshyn, Z. W. (Eds.). (1996). *The Robot's Dilemma Revisited*. Stamford, CT: Ablex Publishers.
- Fox, P. T., Mintun, M. A., Raichle, M. E., Miezin, F. M., Allman, J. M., & Van Essen, D. C. (1986). Mapping human visual cortex with positron emission tomography. *Nature*, 323(6091), 806-809.
- Fraisse, P. (1963). *The Psychology of Time*. New York: Harper & Row.
- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10(1), 126-132.
- Friedman, A. (1978). Memorial comparisons without the "mind's eye." *Journal of Verbal Learning & Verbal Behavior*, 17(4), 427-444.
- Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press (A Bradford Book).
- Gallistel, C. R. (1994). Foraging for brain stimulation: Toward a neurobiology of computation. *Cognition*, 50, 151-170.
- Gardner, M. (1982). *Logic machines and diagrams* (second ed.). Chicago: University of Chicago Press.
- Getzels, J. W., & Jackson, P. W. (1962). *Creativity and intelligence: Explorations with gifted students*. NY: Wiley. (1962). xvii, 293pp.
- Gilden, D., Blake, R., & Hurst, G. (1995). Neural adaptation of imaginary visual motion. *Cognitive Psychology*, 28(1), 1-16.
- Goldenberg, G. (1992). Loss of visual imagery and loss of visual knowledge--a case study. *Neuropsychologia*, 30(12), 1081-1099.
- Goldenberg, G., & Artner, C. (1991). Visual imagery and knowledge about the visual appearance of objects in patients with posterior cerebral artery lesions. *Brain and Cognition*, 15(2), 160-186.
- Goldenberg, G., Mullbacher, W., & Nowak, A. (1995). Imagery without perception--a case study of anosognosia for cortical blindness. *Neuropsychologia*, 33(11), 1373-1382.
- Goodale, M. A., Jacobson, J. S., & Keillor, J. M. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia*, 32(10), 1159-1178.
- Goodman, N. (1968). *Languages of Art*. Indianapolis: Bobbs-Merrill.
- Goryo, K., Robinson, J. O., & Wilson, J. A. (1984). Selective looking and the Mueller-Lyer illusion: The effect of changes in the focus of attention on the Mueller-Lyer illusion. *Perception*, 13(6), 647-654.
- Gregory, R. L. (1968). Visual illusions. *Scientific American*, 219(5), 66-76.
- Grice, H. P. (1975). Logic and Conversation. *Syntax and Semantics*, 3, 41-58.
- Guariglia, C., Padovani, A., Pantano, P., & Pizzamiglio, L. (1993). Unilateral neglect restricted to visual imagery. *Nature*, 364(6434), 235-237.
- Haber, R. N. (1979). Twenty years of haunting eidetic imagery: Where's the ghost? *Behavioral & Brain Sciences*, 2(4), 583-629.
- Haier, R. J., Siegel, B. V., Neuchterlein, K. H., Hazlett, E., Wu, J. C., Paek, J., et al. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *intelligence*, 12, 199-217.
- Hampson, P. J., & Duffy, C. (1984). Verbal and spatial interference effects in congenitally blind and sighted subjects. *Canadian Journal of Psychology*, 38(3), 411-420.
- Hans, M. A. (1974). Imagery and modality in paired-associate learning in the blind. *Bulletin of the Psychonomic Society*, 4(1), 22-24.

- Harper, W. L., Stalnaker, R., & Pearce, G. (Eds.). (1981). *Ifs: Conditionals, belief, decision, chance and time*. Boston, MA: D. Reidel.
- Harris, J. P. (1982). The VVIQ imagery-induced McCollough effects: an alternative analysis. *Percept Psychophys*, 32(3), 290-292.
- Hayes, J. R. (1973). On the Function of Visual Imagery in Elementary Mathematics. In W. G. Chase (Ed.), *Visual Information Processing*. New York: Academic Press.
- Hayes, J. R., & Simon, H. A. (1976). The Understanding Process: Problem Isomorphs. *Cognitive Psychology*, 8, 165-180.
- Heller, M. A., & Kennedy, J. M. (1990). Perspective taking, pictures, and the blind. *Perception & Psychophysics*, 48(5), 459-466.
- Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10(5), 438-443.
- Heywood, C. A., & Zihl, J. (1999). Motion blindness. In G. W. Humphreys (Ed.), *Case studies in the neuropsychology of vision* (pp. 1-16). Hove, England: Psychology Press/Taylor & Francis.
- Hinton, G. E. (1987). The horizontal-vertical delusion. *Perception*, 16(5), 677-680.
- Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception* (pp. 309-331). New York: Holt, Rinehart & Winston.
- Hochberg, J., & Gellman, L. (1977). The effect of landmark features on mental rotation times. *Memory & Cognition*, 5(1), 23-26.
- Hoenig, P. (1972). The effects of eye movements, fixation and figure size on decrement in the Muller-Lyer illusion. *Dissertation Abstracts International*, 33(6-B), 2835.
- Howard, I. P. (1982). *Human Visual Orientation*. New York, NY: John Wiley & Sons.
- Howard, R. J., ffytche, D. H., Barnes, J., McKeefry, D., Ha, Y., Woodruff, P. W., et al. (1998). The functional anatomy of imagining and perceiving colour. *Neuroreport*, 9(6), 1019-1023.
- Humphrey, G. (1951). *Thinking: An introduction to its experimental psychology*. London: Methuen.
- Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review*, 75(6), 550-560.
- Ingle, D. J. (2002). Problems with a "cortical screen" for visual imagery. *Behavioral & Brain Sciences*, 29(2), xxx-xxx.
- Intons-Peterson, M. J. (1983). Imagery paradigms: How vulnerable are they to experimenters' expectations? *Journal of Experimental Psychology: Human Perception & Performance*, 9(3), 394-412.
- Intons-Peterson, M. J., & White, A. R. (1981). Experimenter naivete and imaginal judgments. *Journal of Experimental Psychology: Human Perception & Performance*, 7(4), 833-843.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23, 420-456.
- Ishai, A., & Sagi, D. (1995). Common mechanisms of visual imagery and perception. *Science*, 268(5218), 1772-1774.
- Jamnek, M. (2001). *Mathematical reasoning with diagrams: from intuition to automation*. Chicago, IL: U of Illinois Press (distributor).
- Jankowiak, J., Kinsbourne, M., Shalev, R. S., & Bachman, D. L. (1992). Preserved visual imagery and categorization in a case of associative visual agnosia. *Journal of Cognitive Neuroscience*, 4(2), 119-131.
- Jastrow, J. (1900). *Fact and Fable in Psychology*. New York: Houghton, Mifflin.
- Johnson, R. A. (1980). Sensory images in the absence of sight: Blind versus sighted adolescents. *Perceptual & Motor Skills*, 51(1), 177-178.
- Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *Foundations of cognitive science*. (pp. 469-499). Cambridge, MA, US: The MIT Press.
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, 5(10), 434-442.
- Jonides, J., Kahn, R., & Rozin, P. (1975). Imagery instructions improve memory in blind subjects. *Bulletin of the Psychonomic Society*, 5(5), 424-426.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4), 441-480.
- Kant, E. (1998). *The Critique of Pure Reason (Kritik der reinen Vernunft. English)* (P. Guyer & A. Wood, Trans.). Cambridge, UK: Cambridge University Press.
- Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition*, 79, 89-113.
- Kelso, J. A. S., Cook, E., Olson, M. E., & Epstein, W. (1975). Allocation of attention and the locus of adaptation to displaced vision. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 237-245.

- Kerr, N. H. (1983). The role of vision in "visual imagery" experiments: Evidence from the congenitally blind. *Journal of Experimental Psychology: General*, 112(2), 265-277.
- Klein, G., & Crandall, B. W. (1995). The role of mental simulation in problem solving and decision making. In P. Hancock (Ed.), *Local applications of the ecological approach to human-machine systems, Volume 2: Resources for ecological psychology* (Vol. 2, pp. 324-358). Mahwah, NJ: Lawrence Erlbaum Associates.
- Koenderink, J. J. (1990). *Solid Shape*. Cambridge, MA: MIT Press.
- Kosslyn, S. M. (1978). Measuring the visual angle of the mind's eye. *Cognitive Psychology*, 10, 356-389.
- Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, Mass.: Harvard Univ. Press.
- Kosslyn, S. M. (1981). The Medium and the Message in Mental Imagery: A Theory. *Psychological Review*, 88, 46-66.
- Kosslyn, S. M. (1983). *Ghosts in the Mind's Machine*. New York: Norton.
- Kosslyn, S. M. (1994). *Image and Brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 46-60.
- Kosslyn, S. M., Pascual-Leone, A., Felican, O., Camposano, S., Keenan, J. P., Thompson, W. L., et al. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science*, 284(April 2), 167-170.
- Kosslyn, S. M., Pinker, S., Smith, G., & Shwartz, S. P. (1979). On the demystification of mental imagery. *Behavioral and Brain Science*, 2, 535-581.
- Kosslyn, S. M., Sukel, K. E., & Bly, B. M. (1999). Squinting with the mind's eye: Effects of stimulus resolution on imaginal and perceptual comparisons. *Memory & Cognition*, 27(2), 276-287.
- Kosslyn, S. M., & Sussman, A. L. (1995). Roles of imagery in perception: or, There is no such thing as immaculate perception. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1035-1042). Cambridge, MA: MIT Press.
- Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378(Nov 30), 496-498.
- Kosslyn, S. M., Thomson, W. L., & Ganis, G. (2002). Mental imagery doesn't work like that. *Behavioral & Brain Sciences*, 25(2), xxx.
- Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17(2), 248-294.
- Kowler, E. (1989). Cognitive expectations, not habits, control anticipatory smooth oculomotor pursuit. *Vision Research*, 29, 1049-1057.
- Kowler, E. (1990). The role of visual and cognitive processes in the control of eye movement. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 1-70). Amsterdam: Elsevier Science Publishers.
- Kunen, S., & May, J. G. (1980). Spatial frequency content of visual imagery. *Perception & Psychophysics*, 28(6), 555-559.
- Kunen, S., & May, J. G. (1981). Imagery-induced McCollough effects: Real or imagined? *Perception & Psychophysics*, 30(1), 99-100.
- Landis, T. (2000). Disruption of space perception due to cortical lesions. *Spatial Vision*, 13(2-3), 179-191.
- Levesque, H. (1986). Making believers out of computers. *Artificial Intelligence*, 30, 81-108.
- Leyton, M. (1992). *Symmetry, causality, mind*. Cambridge, MA, USA: Mit Press.
- Luce, R. D., D'Zmura, M., Hoffman, D. D., Iverson, G. J., & Romney, A. K. (Eds.). (1995). *Geometric representations of perceptual phenomena: Papers in honor of Tarow Indow on his 70th birthday* (Vol. 356). Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.
- Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs*, 54(6), Whole No. 248.
- Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, 10(2), 181-187.
- Marmor, G. S., & Zaback, L. A. (1976). Mental rotation by the blind: Does mental rotation depend on visual imagery? *Journal of Experimental Psychology: Human Perception & Performance*, 2(4), 515-521.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

- Marr, D., & Nishihara, H. K. (1976). *Representation and recognition of spatial organization of three-dimensional shapes*. (No. MIT A. I. Memo 377).
- Marr, D., & Nishihara, H. K. (1978). Representation and Recognition of Spatial Organization of Three-Dimensional Shapes. *Proceedings of the Royal Society of London, B*, 200, 269-294.
- Marshall, J. C. (2001). Auditory neglect and right parietal cortex. *Brain*, 124(4), 645-646.
- Mather, J. A., & Lackner, J. R. (1977). Adaptation to visual rearrangement: Role of sensory discordance. *Quarterly Journal of Experimental Psychology*, 29(2), 237-244.
- Mather, J. A., & Lackner, J. R. (1980). Visual tracking of active and passive movements of the hand. *Quarterly Journal of Experimental Psychology*, 32(2), 307-315.
- Mather, J. A., & Lackner, J. R. (1981). Adaptation to visual displacement: Contribution of proprioceptive, visual, and attentional factors. *Perception*, 10(4), 367-374.
- McConkie, G. M., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3), 563-581.
- McDermott, D. (1989). A critique of pure reason. *Computational Intelligence*, 3, 151-160.
- McGlinchey-Berroth, R., Milberg, W. P., Verfaellie, M., & Grande, L. (1996). Semantic processing and orthographic specificity in hemispatial neglect. *Journal of Cognitive Neuroscience*, 8(3), 291-304.
- McKelvie, S. J. (1994). The Vividness of Visual Imagery Questionnaire as a predictor of facial recognition memory performance. *Br J Psychol*, 85(Pt 1), 93-104.
- McKim, R. H. (1978). The Imaginarium: An environment and program for opening the mind's eye. In B. S. Randhawa & W. E. Coffman (Eds.), *Visual Learning, Thinking and Communication* (pp. 61-91). New York: Academic Press.
- Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. (1998). Reopening the mental imagery debate: lessons from functional anatomy. *Neuroimage*, 8(2), 129-139.
- Mellet, E., Tzourio, N., Crivello, F., Joliot, M., Denis, M., & Mazoyer, B. (1996). Functional anatomy of spatial mental imagery generated from verbal instructions. *Journal of Neuroscience*, 16(20), 6504-6512.
- Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: perspectives from cognitive psychology. *Cognition*, 79, 115-134.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Miller, N. (2001). *A Diagrammatic Formal System for Euclidean Geometry*. Unpublished Ph.D., Cornell University, Ithaca, NY.
- Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. New York: Oxford University Press.
- Mitchell, D. B., & Richman, C. L. (1980). Confirmed reservations: Mental travel. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 58-66.
- Moffatt, H. K. (2000). Euler's disk and its finite-time singularity. *Nature*, 404(6780), 833-834.
- Newell, A. (1980). Physical Symbol Systems. *Cognitive Science*, 4(2), 135-183.
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Nichols, S., Stich, S., Leslie, A., & Klein, D. (1996). Varieties of on-line simulation. In P. Carruthers & P. Smith (Eds.), *Theories of theories of mind* (pp. 39-74). Cambridge, UK: Cambridge University Press.
- Nicod, J. (1970). *Geometry and Induction*. Berkeley: Univ. of California Press.
- Nijhawan, R. (1991). Three-dimensional Muller-Lyer Illusion. *Perception & Psychophysics*, 49(9), 333-341.
- Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370(6487), 256-257.
- Norman, D. A. (1988). *The psychology of everyday things*. New York, NY, US: Basic Books, Inc. (1988). xi, 257pp.
- Norman, D. A. (1991). Cognitive artifacts. In J. M. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface* (pp. 17-38). New York, NY, US: Cambridge University Press; Cambridge University Press.
- Ohkuma, Y. (1986). A comparison of image-induced and perceived Mueller-Lyer illusion. *Journal of Mental Imagery*, 10(4), 31-38.
- O'Kane, B. L., Biederman, I., Cooper, E. E., & Nystrom, B. (1997). An account of objects identification confusions. *Journal of Experimental Psychology: Applied*, 3(1), 21-41.
- O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46, 461-488.

- O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, 23(8), 765-768.
- O'Regan, J. K., & Noë, A. (2002). A sensorymotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939-1031.
- Paivio, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Reinhart, and Winston.
- Paivio, A. (1986). *Mental Representations*. New York: Oxford University Press.
- Paivio, A., & te Linde, J. (1980). Symbolic comparisons of objects on color attributes. *Journal of Experimental Psychology: Human Perception & Performance*, 6(4), 652-661.
- Palmer, S. E. (1978). Structural aspects of visual similarity. *Memory and Cognition*, 6, 91-97.
- Patterson, J., & Deffenbacher, K. (1972). Haptic perception of the Mueller-Lyer illusion by the blind. *Perceptual & Motor Skills*, 35(3), 819-824.
- Pavani, F., Ladavas, E., & Driver, J. (2002). Selective deficit of auditory localisation in patients with visuospatial neglect. *Neuropsychologia*, 40(3), 291-301.
- Pentland, A. P. (1986). Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28, 293-331.
- Perky, C. W. (1910). An Experimental study of imagination. *American Journal of Psychology*, 21(3), 422-452.
- Pessoa, L., Thompson, E., & Noë, A. (1998). Finding out about filling in: A guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and Brain Sciences*, 21(6), 723-802.
- Peterson, M. A. (1993). The ambiguity of mental images: insights regarding the structure of shape memory and its function in creativity. In H. Roskos-Ewoldson, M. J. Intons-Peterson & R. E. Anderson (Eds.), *Imagery, creativity, and discovery: A cognitive perspective. Advances in psychology*, Vol. 98. Amsterdam, Netherlands: North Holland/Elsevier Science Publishers.
- Peterson, M. A., Kihlstrom, J. F., Rose, P. M., & Glisky, M. A. (1992). Mental images can be ambiguous: Reconstruals and reference-frame reversals. *Memory and Cognition*, 20(2), 107-123.
- Pinker, S. (1980). Mental imagery and the third dimension. *Journal of Experimental Psychology: General*, 109(3), 354-371.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Pinker, S., Choate, P. A., & Finke, R. A. (1984). Mental extrapolation in patterns constructed from memory. *Memory & Cognition*, 12(3), 207-218.
- Pinker, S., & Finke, R. A. (1980). Emergent two-dimensional patterns in images rotated in depth. *Journal of Experimental Psychology: Human Perception & Performance*, 6(2), 244-264.
- Planck, M. (1933). *Where is Science Going?* (J. Murphy, Trans.). London: Allen & Unwin.
- Podgorny, P., & Shepard, R. N. (1978). Functional representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception and Performance*, 4(1), 21-35.
- Pointcaré, H. (1963). Why space has three dimensions. In *Mathematics and Science: Last Essays*. New York: Dover.
- Predebon, J., & Wenderoth, P. (1985). Imagined stimuli: Imaginary effects? *Bulletin of the Psychonomic Society*, 23(3), 215-216.
- Prinz, J. J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W. (1973). What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psychological Bulletin*, 80, 1-24.
- Pylyshyn, Z. W. (1978). Imagery and Artificial Intelligence. In C. W. Savage (Ed.), *Perception and Cognition: Issues in the Foundations of Psychology* (Vol. 9). Minneapolis: Univ. of Minnesota Press.
- Pylyshyn, Z. W. (1979a). Do Mental Events Have Durations? *Behavioral and Brain Sciences*, 2(2), 277-278.
- Pylyshyn, Z. W. (1979b). The Rate of 'Mental Rotation' of Images: A Test of a Holistic Analogue Hypothesis. *Memory and Cognition*, 7, 19-28.
- Pylyshyn, Z. W. (1979c). Validating Computational Models: A Critique of Anderson's Indeterminacy of Representation Claim. *Psychological Review*, 86(4), 383-394.
- Pylyshyn, Z. W. (1980). Cognitive Representation and the Process-Architecture Distinction. *Behavioral and Brain Sciences*, 3(1), 154-169.
- Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 88, 16-45.

- Pylyshyn, Z. W. (1984a). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W. (1984b). Plasticity and Invariance in Cognitive Development. In J. M. a. R. Fox (Ed.), *Neonate Cognition: Beyond the Blooming, Buzzing Confusion*. Hillsdale, N.J.: Erlbaum.
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32, 65-97.
- Pylyshyn, Z. W. (1991a). The role of cognitive architectures in theories of cognition. In K. VanLehn (Ed.), *Architectures for Intelligence*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pylyshyn, Z. W. (1991b). Rules and Representation: Chomsky and representational realism. In A. Kashir (Ed.), *The Chomskian Turn*. Oxford: Basil Blackwell Limited.
- Pylyshyn, Z. W. (1996). The study of cognitive architecture. In D. Steier & T. Mitchell (Eds.), *Mind Matters: Contributions to Cognitive Science in honor of Allen Newell*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3), 341-423.
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1/2), 127-158.
- Pylyshyn, Z. W. (2002). Mental Imagery: In search of a theory. *Behavioral and Brain Sciences*, 25(2), xxx-xxx.
- Pylyshyn, Z. W. (Ed.). (1987). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. Norwood, NJ: Ablex Publishing.
- Pylyshyn, Z. W., & Cohen, J. (1999). *Imagined extrapolation of uniform motion is not continuous*. Paper presented at the Annual Conference of the Association for Research in Vision and Ophthalmology, May 1999., Ft. Lauderdale, FL.
- Ranney, M. (1989). Internally represented forces may be cognitively penetrable: Comment on Freyd, Pantzer, and Cheng (1988). *Journal of Experimental Psychology: General*, 118(4), 399-402.
- Reed, S. K., Hock, H. S., & Lockhead, G. R. (1983). Tacit knowledge and the effect of pattern configuration on mental scanning. *Memory and Cognition*, 11, 137-143.
- Reisberg, D., & Chambers, D. (1991). Neither pictures nor propositions: What can we learn from a mental image? *Canadian Journal of Psychology*, 45(3), 336-352.
- Reisberg, D., & Morris, A. (1985). Images contain what the imager put there: A nonreplication of illusions in imagery. *Bulletin of the Psychonomic Society*, 23(6), 493-496.
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, 7, 17-42.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368-373.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, 7, 127-145.
- Richman, C. L., Mitchell, D. B., & Reznick, J. S. (1979). Mental travel: Some reservations. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 13-18.
- Richter, W., Somorjai, R., Summers, R., Jarmasz, M., Menon, R. S., Gati, J. S., et al. (2000). Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience*, 12(2), 310-320.
- Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 1157-1165.
- Rieser, J. J., Guth, D. A., & Hill, E. W. (1986). Sensitivity to perspective structure while walking without vision. *Perception*, 15, 173-188.
- Rode, G., Rossetti, Y., & Bission, D. (2001). Prism adaptation improves representational neglect. *Neuropsychologia*, 39(11), 1250-1254.
- Roland, P. E., & Gulyas, B. (1994a). Beyond 'pet' methodologies to converging evidence: Reply. *Trends in Neurosciences*, 17(12), 515-516.
- Roland, P. E., & Gulyas, B. (1994b). Visual imagery and visual representation. *Trends in Neurosciences*, 17(7), 281-287.
- Roland, P. E., & Gulyas, B. (1995). Visual memory, visual imagery, and visual recognition of large field patterns by the human brain: Functional anatomy by positron emission tomography. *Cerebral Cortex*, 5(1), 79-93.
- Russell, B. (1918/1985). The Philosophy of Logical Atomism. In D. F. Pears (Ed.), *The Philosophy of Logical Atomism* (pp. 35-155). Lasalle: Open Court.

- Samson, D., Pillon, A., & De Wilde, V. (1998). Impaired knowledge of visual and non-visual attributes in a patient with a semantic impairment for living entities: A case of a true category-specific deficit. *Neurocase: Case Studies in Neuropsychology, Neuropsychiatry, & Behavioural Neurology*, 4(4-5), 273-290.
- Sarter, M., Berntson, G. G., & Cacioppo, J. T. (1996). Brain imaging and cognitive neuroscience: Toward strong inference in attributing function to structure. *American Psychologist*, 51(1), 13-21.
- Schweinberger, S. R., & Stief, V. (2001). Implicit perception in patients with visual neglect: Lexical specificity in repetition priming. *Neuropsychologia*, 39(4), 420-429.
- Searle, J. R. (1990). Consciousness, explanatory inversion, and cognitive science. *Behavioral & Brain Sciences*, 13(4), 585-642.
- Segal, S. J., & Fusella, V. (1969). Effects of imaging on signal-to-noise ratio, with varying signal conditions. *British Journal of Psychology*, 60(4), 459-464.
- Segal, S. J., & Fusella, V. (1970). Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology*, 83(3), 458-464.
- Sergent, J. (1994). Brain-imaging studies of cognitive functions. *Trends in Neurosciences*, 17, 2 21-227.
- Servos, P., & Goodale, M. A. (1995). Preserved visual imagery in visual form agnosia. *Neuropsychologia*(Nov), 1383-1394.
- Sheingold, K., & Tenney, Y. J. (1982). Memory for a salient childhood event. In U. Neisser (Ed.), *Memory Observed* (pp. 201-212). San Francisco, CA: W.H. Freeman and Co.
- Shepard, R. N. (1978a). Externalization of mental images and the act of creation. In B. S. Randhawa & W. E. Coffman (Eds.), *Visual Learning, Thinking and Communication* (pp. 133-189). New York: Academic Press.
- Shepard, R. N. (1978b). The Mental Image. *American Psychologist*, 33, 125-137.
- Shepard, R. N., & Chipman, S. (1970). Second-Order Isomorphism of Internal Representations: Shapes of States. *Cognitive Psychology*, 1, 1-17.
- Shepard, R. N., & Feng, C. (1972). A Chronometric Study of Mental Paper Folding. *Cognitive Psychology*, 3, 228-243.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science*, 171, 701-703.
- Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, 101(2), 357-361.
- Shioiri, S., Cavanagh, P., Miyamoto, T., & Yaguchi, H. (2000). Tracking the aparent location of targets in interpolated motion. *Vision Research*, 40, 1365-1376.
- Shuren, J. E., Brott, T. G., Schefft, B. K., & Houston, W. (1996). Preserved color imagery in an achromatopsic. *Neuropsychologia*, 34(6), 485-489.
- Silbersweig, D. A., & Stern, E. (1998). Towards a functional neuroanatomy of conscious perception and its modulation by volition: implications of human auditory neuroimaging studies. *Philos Trans R Soc Lond B Biol Sci*, 353(1377), 1883-1888.
- Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, Mass.: MIT Press.
- Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science*, 7(5), 301-305.
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1, 261-267.
- Slezak, P. (1991). *Can images be rotated and inspected? A test of the pictorial medium theory*. Paper presented at the Thirteenth Annual meeting of the Cognitive Science Society.
- Slezak, P. (1992). *When can images be reinterpreted: Non-chronometric tests of pictorialism*. Paper presented at the Fourteenth Conference of the Cognitive Science Society.
- Slezak, P. (1995). The 'philosophical' case against visual imagery. In P. Slezak, T. Caelli & R. Clark (Eds.), *Perspective on Cognitive Science: Theories, Experiments and Foundations* (pp. 237-271). Stamford, CT: Ablex.
- Sloman, A. (1971). Interactions between philosophy and artificial intelligence: The role of intuition and non-logical reasoning in intelligence. *Artificial Intelligence*, 2, 209-225.
- Sperber, D., & Wilson, D. (1998). The mapping between the mental and the public lexicon. In P. Carruthers & J. Boucher (Eds.), *Thought and Language* (pp. 184-200). Cambridge, UK: Cambridge University Press.
- Squire, L. R., & Slater, P. C. (1975). Forgetting in very long-term memory as assessed by an improved questionnaire technique. *Journal of Experimental Psychology: Human Perception and Performance*, 104, 50-54.
- Steinbach, M. J. (1976). Pursuing the perceptual rather than the retinal stimulus. *Vision research*, 16, 1371-1376.
- Stoerig, P. (1996). Varieties of vision: From blind responses to conscious recognition. *Trends in Neurosciences*, 19(9), 401-406.

- Stromeyer, C. F., & Psotka, J. (1970). The detailed texture of eidetic images. *Nature*, 225(230), 346-349.
- Suppes, P. (1995). Some foundational problems in the theory of visual space. In R. D. Luce & M. e. a. D'Zmura (Eds.), *Geometric representations of perceptual phenomena: Papers in honor of Tarow Indow on his 70th birthday* (pp. 37-45). Mahwah, NJ: Lawrence Erlbaum Associates.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning & Verbal Behavior*, 18(6), 645-659.
- Talmy, L. (2000). *Toward a cognitive semantics, Vol. 1: Concept structuring systems*: Cambridge, MA, US: The MIT Press. (2000). viii, 565pp.
- Tipper, S. P., & Behrmann, M. (1996). Object-centered not scene-based visual neglect. *Journal of Experimental Psychology: Human Perception and Performance*, 22(5), 1261-1278.
- Tlauka, M., & McKenna, F. P. (1998). Mental imagery yields stimulus-response compatibility. *Acta Psychologica*, 67-79.
- Todd, J. T., Tittle, J. S., & Norman, J. F. (1995). Distortions of three-dimensional space in the perceptual analysis of motion and stereo. *Perception*, 24(1), 75-86.
- Tootell, R. B., Silverman, M. S., Switkes, E., & de Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218(4575), 902-904.
- Treisman, A., & Schmidt, H. (1982). Illusory Conjunctions in the Perception of Objects. *Cognitive Psychology*, 14(1), 107-141.
- Trojano, L., & Grossi, D. (1994). A critical review of mental imagery defects. *Brain & Cognition*, 24(2), 213-243.
- Tsotsos, J. K. (1988). How does human vision beat the computational complexity of visual perception. In Z. W. Pylyshyn (Ed.), *Computational Processes in Human Vision: An interdisciplinary perspective* (pp. 286-340). Norwood, NJ: Ablex Publishing.
- Tufte, E. R. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.
- Tye, M. (1991). *The Imagery Debate*. Cambridge, MA: MIT Press.
- Uhlarik, J. J. (1973). Role of cognitive factors on adaptation to prismatic displacement. *Journal of Experimental psychology*, 98, 223-232.
- Ullman, S. (1984). Visual routines. *Cognition*, 18, 97-159.
- Virsu, V. (1971). Tendencies to eye movement, and misperception of curvature, direction, and length. *Perception & Psychophysics*, 9(1-B), 65-72.
- Vuilleumier, P., & Rafal, R. (1999). "Both" means more than "two": Localizing and counting in patients with visuospatial neglect. *Nature Neuroscience*, 2(9), 783-784.
- Wallace, B. (1984a). Apparent equivalence between perception and imagery in the production of various visual illusions. *Memory & Cognition*, 12(2), 156-162.
- Wallace, B. (1984b). Creation of the horizontal-vertical illusion through imagery. *Bulletin of the Psychonomic Society*, 22(1), 9-11.
- Wallace, B. (1991). Imaging ability and performance in a proofreading task. *Journal of Mental Imagery*, 15(3-4), 177-188.
- Watanabe, K., & Shimojo, S. (1998). Attentional modulation in perception of visual motion events. *Perception*, 27(9), 1041-1054.
- Watson, D. G., & Humphreys, G. W. (1997). Visual marking: prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review*, 104(1), 90-122.
- Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, 68(1), 77-94.
- Wittgenstein, L. (1953). *Philosophical Investigations [Philosophische Untersuchungen]*. Oxford: Blackwell.
- Wyatt, H. J., & Pola, J. (1979). The role of perceived motion in smooth pursuit eye movements. *Vision Research*, 19, 613-618.
- Yantis, S., & Jones, E. (1991). Mechanisms of attentional selection: temporally modulated priority tags. *Perception and Psychophysics*, 50(2), 166-178.
- Zhou, H., & May, J. G. (1993). Effects of spatial filtering and lack of effects of visual imagery on pattern-contingent color aftereffects. *Perception and Psychophysics*, 53, 145-149.
- Zimler, J., & Keenan, J. M. (1983). Imagery in the congenitally blind: How visual are visual images? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9(2), 269-282.