IS THE IMAGERY DEBATE OVER? IF SO, WHAT WAS IT ABOUT?

Zenon Pylyshyn
Rutgers Center for Cognitive Science
Rutgers University, New Brunswick, NJ

Background

Jacques Mehler was notoriously charitable in embracing a diversity of approaches to science and to the use of many different methodologies. One place where his ecumenism brought the two of us into disagreement is when the evidence of brain imaging was cited in support of different psychological doctrines, such as the picture-theory of mental imagery. Jacques remained steadfast in his faith in the ability of neuroscience data (where the main source of evidence has been from clinical neurology and neuro-imaging) to choose among different psychological positions. I personally have seen little reason for this optimism so Jacques and I frequently found ourselves disagreeing on this issue, though I should add that we rarely disagreed on substantive issues on which we both had views. This particular bone of contention, however, kept us busy at parties and during the many commutes between New York and New Jersey, where Jacques was a frequent visitor at the Rutgers Center for Cognitive Science. Now that I am in a position where he is a captive audience it seems an opportune time to raise the question again.

I don't intend to make a general point about sources of evidence. It may even be, as Jacques has frequently said, that we have sucked dry the well of reaction-time data (at least in psycholinguistics), and that it is time to look elsewhere. It may even be that the evidence from PET and fMRI will tell us things we did not already know – who can foresee how it will turn out? But one thing I can say with some confidence is that if we do not have a clear statement of the question we are trying to answer, or of an internally coherent hypothesis, neither reaction time nor PET nor fMRI nor rTMS will move us forward. So long as we are engaged in a debate in which the basic claims are muddled or stated in terms of metaphors that permit us to freely back out when the literal interpretation begins to look untenable, then we will not settle our disagreements by merely deploying more hi-tech equipment. And this, I suggest, is exactly what is happening in the so-called "imagery debate", notwithstanding claims that the debate has been "resolved" (Kosslyn, 1994).

The historical background to the imagery debate

This essay is in part about the "debate" concerning the nature of mental imagery. Questions about the nature of conscious mental states have a very long history in our attempt to understand the mind. Pre-theoretically it has been apparent that we think either in words or in pictures. In the last 40 years this idea has been worked into theories of mental states within the information-processing or computational view of mental processes. The so-called "dual code" view has been extremely influential in psychology since about 1970, due largely to the early work of Allan Paivio (Paivio, 1971) (for an early critique of this work, see Pylyshyn, 1973). Shortly after this renaissance in interest in mental imagery, the emphasis turned from the study of learning and the appeal to imagery as a intervening variable to an attempt to work out the nature of mental images themselves (this work is summarized in Kosslyn, 1980) (for a critique of this later work, see Pylyshyn, 1981). Led by the influential work of Stephen Kosslyn, researchers investigated the structure of mental images, including their metrical properties. For example, images seemed to actually *have* distance, since it took longer to scan greater distance in an image; they seemed to *have* size inasmuch as it took longer to report small features in a small image than in a large one; the "mind's eye" that inspects images also seemed to have horizontal and vertical limits and its resolution fell off with

eccentricity much as that of the real eye. In addition it appears that we can manipulate images much the way we can manipulate physical objects; we can rotate mental images (in three dimensions), we can fold them and watch what happens, we can draw things on them or superimpose other images on them or on our percepts, and so on. Such abilities suggested to researchers that images must have a special form or underlying instantiation in the brain and many researchers proposed that images differ from other forms of representation (presumably "verbal" representations) in that they have spatial properties, are displayed in the brain, and represent by virtue of "depicting" or by virtue of resembling what they represent, rather than by virtue of describing their target scenes.

Throughout these developments I have maintained that we are under a collective illusion (a "grande illusion", to use the French phrase). The illusion is that when we experience "seeing an image with the mind's eye" we are actually inspecting a mental state, a structure that can play a role in an information processing account of mental activity. I argued that what was going on in these studies is that subjects were being asked, in effect, what it would be like to *see* certain things happening (a scene being scanned by attention, looking for a small detail in a scene, or watching an object being rotated). In my critique, I suggested that what was happening in these studies was that people were using what they know about thre world to *simulate* certain observable aspects of the sequence of events that would have unfolded in the situation being studied. In other words, I claimed that the experiments were revealing what subjects believed about what would happen if they were looking at a certain scene and not the inherent nature an imagery medium or mechanism.

Such claims and counter claims went on for two decades. Then in the last few years a new source of evidence was introduced which many people, like Stephen Kosslyn took to provide (in the words of the subtitle of Kosslyn's influential book Kosslyn, 1994), "The resolution of the imagery debate". Many investigators were persuaded that here, finally, was evidence that was direct and unambiguous and proved that there were images in the brain – actual displays realized as patterns of activity in the visual cortex. What are we to make of these new results, which have persuaded a large number of researchers of the basic correctness of the picture-theory? Is it true that we now have concrete evidence about the nature of mental images when previously we had only indirect and ambiguous behavioral evidence?

Many things to many people

In discussing this question, I will begin by laying out what I think is the state of the "debate". Later I will come back to the question of whether neuropsychological evidence has made (or is likely to make) any difference. What is the problem? To put it as bluntly as possible, the problem over the question of the nature of mental imagery is just that some people find the experience of seeing a picture-in-the-head completely compelling while others find it irrelevant as the basis for a theory of what is actually going on in the mind when we entertain mental images. Beyond that there is no useful *general* debate, only arguments about the coherence and validity of certain quite specific proposals, and about what morals can be drawn from particular experiments. This is, in the end, not about metaphysics. It is a scientific disagreement, but one in which there are very many different claims falling under the same umbrella term "the imagery debate" and they may have little or nothing in common, other than sharing the same gut reaction to the picture metaphor and to the question of what to make of the phenomenology of imagery.

Among the many things involved in this debate are a variety of substantive disagreements.

- A disagreement about whether the form of representation underlying mental imagery is in some way special, and if so, in what way it is special – whether it is special because it uses distinct mechanisms specific to imagery or whether it is special because it deals with information about how things look (i.e., because of what imaginal representations are about or their subject matter);

- A disagreement about whether images are "depictive" or whatever the opposite is (perhaps "descriptive");
- A disagreement about whether or not mental imagery "involves the visual system", which itself raises the question of what exactly *is* the visual system and in what way it may be involved;
- A disagreement about whether certain phenomena observed during episodes of mental imagery are due to the fact that the brain evolved in a particular way resulting in a "natural harmony" between the way things unfold in one's imagination and the way they unfold in the world or in one's perception of the world (what Shepard, 1975, has called "second order isomorphism").
- A disagreement about whether some of the phenomena observed during episodes of mental imagery arise because (1) people are reasoning from what they know about the situation being imagined, and are simulating what they believe would have happened if a real event were being observed, or because (2) special image-specific mechanisms are deployed when one reasons using mental images. This is the disagreement that I wrote about in (Pylyshyn, 1981) and, I believe, remains one of the main questions about mental imagery.

One of the things that makes this debate both ironic and ill-posed is that it is hard to disagree with most of the picture theory[1] views in this discussion, since there is *something* right about the claims. It is true that when we solve problems in which the geometry of a display plays a roll we usually do so by imagining the figure, and that when we do imagine the figure we are able to do things we could not do if we were to approach the problem symbolically – say by thinking in words. This much is not in dispute. What is in dispute is what is going on in our head when we engage in the sort of activity we call "imaging", and in particular what an adequate theory of this process will need to postulate as the bearer of the imaginal information. This is not a case of believing that images do not exist or are "epiphenomenal". It is a question of whether theories of mental imagery that posit 2D displays or "depictive representations" are empirically correct, or perhaps even coherent. In every case I have looked at, hypothesizing pictures or depictions does not provide any explanatory advantage over what I will call the "null hypothesis" that image content is represented as symbolic expressions (see below), even though it may feel more comfortable because it comports with one's subjective impression of what is going on. I, for one, get very nervous when I find a theory in psychology making claims that are consonant with how it looks to me from the inside – I know of too many examples where how it feels on the inside is exactly the wrong kind of theory to have. Things rarely are how they seem in any mature science, and this is especially true in a nascent science like psychology or psychobiology!

Intrinsic and extrinsic constraints

The basic dilemma is that while the following two claims may both seem to be true, they are incompatible – at least in their naïve form:

(1) Since the mental images we have are of our own making we can make them have whatever properties we wish, and we can make them behave in any way we choose. Having an image is like having a thought – it seems as though we can think any thought there is to think. Consequently *which* image property or thought we have depends on what we want to do, and this generally depends on what we believe.

[1] For the sake of expository simplicity I will refer to the set of ideas motivated by the lure of the phenomenology of pictures-in-the-head as *picture theories*, recognizing that this may be a very heterogeneous set of theories.

(2) Mental images appear to have certain inherent properties that allow them to parallel many aspects of the perceived world. For example images do not seem to unfold the way thoughts do, following principles of inference (including heuristic rules), but in a way that directly reflects how we perceive the world. An image *looks* a certain way to us, therefore we "see" things in them independent of our explicit knowledge about what we are imaging. If we imagine a geometrical figure, such as a parallelogram, and imagine drawing its diagonals we can *see* that one of the diagonals is longer than the other and yet that the two cross at their mutual midpoints, and we can see that this is so apparently without having to infer it from our knowledge of geometry. Similarly, when we imagine a dynamic situation or event unfold in our mind (such as a baseball being hit), the imagined event behaves in ways that appear to be at least partially outside our voluntary control, and maybe even outside our intellectual capacity to calculate.

There are anecdotes to illustrate each of these two perspectives and the opposing factions in the "imagery debate" generally emphasize one or the other of the above claims. The picture-theorists argues that if you have a visual image of an object, then you have no choice but to imagine it from a particular viewpoint and having a particular shape and orientation, etc. Similarly, if you imagine an object as small in size it follows from the inherent nature of mental imagery that it will be harder to "see" and therefore to report small visual details than if you imagined it large; or that if you focus your attention on a place in your imagine and then try to report a property that is far away, it will take longer than if you attempted to report a property that was nearer to where you are focused. The critics of picture theories argue equally cogently that you can just as easily imagine an object without imagining it as having any particular properties, that there is no reason (other than the implicit requirement of the instruction to "imagine something small") why you can't imagine something as small but highly detailed and therefore not take longer to report its visible details, or to imagine switching your attention from one place on an image to another in a time independent of how far away the two places are, *as long as you are not attempting to simulate what would happen if you were looking at the real object* being represented or are not attempting to simulate a situation in which you believe it would take more time to get from one place to another if the places were further apart (in fact I reported the results of several experiments showing that this is exactly what happens Pylyshyn, 1981). The imagery-as-general-reasoning adherents can point to many examples where the way a dynamic image unfolds is clearly under our voluntary control. For example, imagine sugar being poured into a glass full of water (does it overflow?), or imagine yellow and blue transparent colored filters being moved together so they overlap (what color do you see where they overlap?). The answer you give to these questions clearly depends on what you know about the physics of solutions and what you know (or remember) about the psychophysics of color mixing.

It is easy enough to come up with examples that go either way; some empirical phenomena appear to be a consequence of inherent properties of an image representation and others appear to arise because of what you know or believe (perhaps falsely) about how the visually perceived world works. The substantive empirical question is: Which properties are inherent properties of the imagery system (or the medium or mechanisms of mental imagery) and which are properties that the person doing the imaging creates in order to simulate what he or she believes would be true of a real situation corresponding to the one being imagined.

I prefer to put this opposition slightly differently: Some properties of the image or of the imagined situation are cognitively penetrable by knowledge and beliefs (many of which are held tacitly and cannot be reported on demand, as is the case with knowledge of grammar or of many

social conventions).[2] Other properties may be due to intrinsic causes of various sorts, to the architecture of mind. The inherent nature of mental images might be one of the determinants of certain experimental phenomena reported in the literature, but so might the way in which you have learned certain things you know and the way in which you have organized this knowledge (which may have nothing to do with properties of imagery itself). For example, you have learned the alphabet in serial order so in order to tell whether *L* comes after *H* you may have to go through the list, and in order to tell whether certain things happen when you fold a certain paper template to make a cube, you may have to go through a sequence of asking what happens after individual folds (as in the study by Shepard & Feng, 1972, in which it was found that when observers used their imagination to judge whether two arrows in a paper template would touch when folded, it took them longer under just those conditions when it would have taken more folds to actually fold the template). Problems are *generally* solved by the application of a sequence of individual operations so this in itself says nothing special about mental imagery. It's true that in order to recall how many windows there are in your living room you may have to *count* them because the numerical fact is not stored as such. But this has nothing to do with the use of imagery per se, any more than that fact that in order to recall the second line of a poem you need to recall the first line, or that in order to tell how many words it has in it you need to recall the line and count them. There are also many reasons why you might observe certain reliable patterns whenever the subjective experience of "seeing with the mind's eye" occurs. The burden of proof must fall on those who wish to argue in favor of some particular *special mechanism* to show that it is at least unlikely that the general mechanism, that we know exists because it has to be used in non-imagery contexts, will not do.

The reason that there has been so much talk (by me and others) about the representations underlying mental imagery being *propositional* is that there are very good reasons for thinking that much of cognition depends on a l*anguage of thought* (Fodor, 1975; Fodor & Pylyshyn, 1988; Pylyshyn, 1984). For example, propositions, or more correctly, language-like tree-structured symbolic encodings, are the only form of representation that we know that can take advantage of mechanical reasoning mechanisms, such as computers, and they are also the only ones we know that exhibit the properties of compositionality, productivity and systematicity that are essential characteristics of at least human thought (see Fodor & Pylyshyn, 1988). Although that does not entail that mental images are propositions, the propositional proposal serves as the natural *null hypothesis* against which to compare any proposal for a special for of representation for mental imagery. It's not that the idea of images having the form of a set of sentences in some mental calculus is a particularly attractive or natural alternative, but it is the only one so far proposed that is not seriously flawed.

Here is the crux of the problem that picture-theories must face if they are to provide full explanatory accounts of the phenomena. They must show that the relevant empirical phenomena, whether it is the increased time it takes to switch attention to more distant places in an image or the increased time it takes to report details from smaller images, follow from *the very nature of*

---

[2] There has been a great deal of misunderstanding about the notion of tacit knowledge (or tacit beliefs). It is a perfectly general property of knowledge that it can be *tacit,* or not consciously available (Fodor, 1968). We can have knowledge about various aspects of the social and physical world that (1) qualifies as real knowledge, in the sense that it can be shown to enter into general inferences and to account for a wide range of behaviors, and (2) can't be used in answering a direct question. Our knowledge of intuitive physics is of this sort. We have well-developed intuitions about how objects will fall, bounce and accelerate, even though we very often cannot answer abstract questions about it (and indeed we often hold explicit beliefs that are wrong and contradict the way we act towards these objects). Even the knowledge of something that is explicitly taught, such as the procedure for adding numbers is tacit – the rules we might give are generally not the ones that play the causal role in our numerical calculations (VanLehn, 1990).

mental images or of the mechanisms involved in their use. In other words it must be that these phenomena reveal a constraint attributable to the intrinsic nature of the image, to its form or neural implementation, or to the mechanisms that it uses – rather than to some other extrinsic constraint arising from the knowledge that the subject possesses, or from the way this knowledge is structured, or from the subject's goals or understanding of the task. If, in order to account for the regularities, one has to appeal to something other than the inherent constraints of the imagery system then, however one might like the picture-theory as a description of what is going on in the mind, it will not serve as an *explanation*. That is because it is the extrinsic factors that are doing the work and they can equally be applied to *any* form of representation, including one that is propositional. So, for example, if a picture-theory is to explain why it takes longer to switch attention between more distant places in an image one must show that this is *required* by the imagery mechanism or medium or format –because of its very nature or causal structure (e.g., because of the physical laws that apply). Otherwise the appeal to imagery carries no explanatory weight. *Any* form of representation can give the same result merely by adding the stipulation that switching attention between representations of more distant places requires more time (during which, for example, one might entertain thoughts of the form "now it is here", "now it is there" and so on, providing a sequence of thoughts that simulate what might happen if one were looking at a scene). So if you can show empirically that it is unlikely that the properties you observe are due to inherent properties of the image, as opposed to properties of the world envisioned, the reason for preferring the picture-theory would evaporate.

Although this is a simple point it turns out to be one that people have a great deal of difficulty in grasping, so I will try to provide an additional example. Consider the proposal that images need not be literally written on a two-dimensional surface, but rather may be implemented in a *functional space* such as a matrix data structure in a computer. Notice that physical laws do not apply to a functional space. There is nothing about a matrix data structure that *requires* that in order to get from one cell to another you have to pass through intervening cells. In the matrix a "more distant cell" is not actually further away so no physical law requires that it take more time: In fact in a computer one can get from any cell to any other cell in constant time. So if we do require that the process pass through certain other cells, then we are appealing to a constraint extrinsic to the nature of the matrix or "functional space". Of course one might find it *natural* to assume that in order to go from one cell to another the locus of attention must go through intervening ones. But the intervening cells are not in any relevant sense located *between* two other cells except by virtue of the fact that we usually picture matrices as two dimensional tables or surfaces. In a computer we *can* (though we don't have to – except again by extrinsic stipulation) go from one cell to another by applying a successor function to the coordinates (which are technically just ordered names). Thus we *can* require that in going from one cell to another we have to step through the cells that fall between the two, where the relation "between" is defined in terms of the ordering of their names. Thus we can ensure that more such cells are visited when the distance being represented is greater. But this requirement does not follow from the intrinsic nature of a matrix data structure, it is an *added* or extrinsic requirement, *and thus could be imposed equally on any form of representation, including a non-pictorial one*. All that is required is (1) that there be some way of representing potential (or empty) locations and of identifying them as being "in between," and (2) that in accessing places in the representation, those marked as "in between" have to be visited in getting from the representation of one place to the representation of another place. As regards requirement (1), it can be met by any form of representation, including a propositional or symbolic one, so long as we have names for places – which is what Cartesian coordinates (or, for that matter, any compressed form of encoding of pictures such as GIF or JPEG) give us.

The test of whether any particular phenomenon is attributable to the intrinsic nature of images or to tacit knowledge is to see whether the observations in question change in a rationally

comprehensible way if we change the relevant knowledge, beliefs or goals. Take, for example, the robust finding that the time it takes to switch from examining one place on an image to examining another increases linearly with the distance being imagined, a result consistently interpreted to show that images have metrical properties like distance. One can ask whether this time-distance relation arises from an intrinsic property of an image or from the observers' understanding that they are to simulate what happens when looking at a particular display. It is clear that observers can scan an image at a particular speed, or they can scan it at a different speed, or they can simply not scan it at all when switching their attention from one place to another. In our own research, we showed that when observers are given a task that requires focusing on distinct places but that does not emphasize imagining getting from one place to another, the scanning phenomenon vanishes (Pylyshyn, 1981). As in the original scanning experiments, the setup always involved focusing on a place on a mental map and then focusing at another place on the map. But in one experiment the ostensible task in focusing on the second place was to judge the direction of the first place from it (by naming a clock direction). In this and other similar tasks[3] there is no effect of image distance on the time to switch attention between places.

I might note in passing that it is not by any means obvious that people do, in fact represent a succession of empty spaces in scanning studies or in any dynamic visualization. We have obtained some preliminary data (Pylyshyn & Cohen, 1999) suggesting that when we imagine continuously scanning a space between two locations we do not actually traverse a succession of intermediate places unless there are visible features at those locations. When there are such features, it appears that we carry out a sequence of time-to-contact computations to selected visible features along the scan path. Thus it may well be that scanning involves computing a series of times between intermediate visible features and *simulating* the scanning by waiting out the appropriate amount of time for each transition.[4] Note also that while requirement (2) may seem unnatural and unmotivated when applied to a list of sentences, *it is exactly as well-motivated, no more and no less*, as it is when applied to a matrix or other "functional space". In both cases the constraint functions as a free empirical parameter, filled in solely to match the data for the particular case. The same is not true, of course, when the space is a real physical space rather than a "functional" space since there is, after all, a physical law relating time, distance and speed, which applies to real space but not to "functional space". This is why there has been so much interest in finding a real spatial representation of images, a pursuit to which I now turn.

---

[3] Imagine a map with a single illuminated light which goes off at a specified time. Imagine that whenever a light goes off at one place another simultaneously goes on at another place. Now indicate *when you see* the next light come on in your image. In such an experiment there is no increase in time to report seeing the light coming on as a function of the distance between lights.

[4] It should be mentioned in passing that there is an important caveat to be made here concerning cases in which imagery is studied by having subjects "project" their image onto a visible scene, which includes the vast majority of mental imagery studies. In these cases there is a real space, complete with properties of rigidity and stability, in which all the Euclidean axioms hold. If subjects are able to think demonstrative thoughts such as "this" and "that" (as I have elsewhere claimed they can Pylyshyn, in press) and to bind imagined properties to those visible features, then there is a real literal sense in which the spatial properties of the scene are inherited by the image. For example, in such a situation subjects can literally move their eyes to real places where they think of certain imagined properties as being located. Thus it is easy to see how one might get a distance-scanning effect as well as other spatial effects (like *noticing* that one of the imagined objects lies between two other imagined objects). Many imagery results fall out of such a use of a real spatial backdrop (Pylyshyn, 1989; Pylyshyn, in press).

Neuropsychological Evidence and the "new stage" of the debate

The arguments I have sketched (when fleshed out in detail, as I have done elsewhere Pylyshyn, 1981; Pylyshyn, submitted) should make it clear that a picture theory that appeals to *inherent properties* of a "picture-like" 2D display is no better at explaining the results of mental imagery studies than is the "null hypothesis" that the content of our images is encoded in some symbolic form which serve as the basis for inferences and for simulating various aspects of what it would be like to see some situation unfold (including the relative times taken for different tasks). The basic problem is that the phenomena that have attracted people to the picture theory (phenomena such as mental scanning or the effect of image size on reaction times for detecting features) appear to be cognitively penetrable and thus cannot be attributed to the nature of the image itself – to how it is instantiated in brain tissue – as opposed to what people know or infer or assume would happen in the real referent situation. Any attempt to minimize this difficulty, say by postulating that images are only "functionally" like 2D pictures, is powerless to explain the phenomena at all since functional spaces are whatever we want them to be and are thus devoid of explanatory force. But what about the literal conception of images as real 2D displays in the brain? This is the view that is now popular in certain neuropsychology circles and has led to what (Kosslyn, 1994) has described as the "third phase" of the imagery debate – the view that the evidence of neuroscience can reveal the "display" or "depictive" nature of mental images. So where does such evidence place the current state of understanding of mental imagery?

Neuropsychological evidence has been cited in favor of a weak and a strong thesis with little care taken to distinguish them. The weak thesis is that mental imagery in some way involves the visual system. This claim is weak because nobody would be surprised if some parts of visual processing overlaps with virtually any cognitive activity – much depends on what one takes to be "the visual system." (for more on this question see Pylyshyn, 1999). The strong claim is that not only is the visual system involved, but the input to this system is a spatially laid out as a "picture-like" pattern of activity. Yet the evidence cited in favor of the weak thesis that imagery involves the visual system is often also taken (sometimes implicitly and sometimes explicitly) to support the stronger thesis, that images are structurally different from other forms of thought because they are laid out spatially the way pictures are, and therefore that they are not descriptive but *depictive* (whatever exactly that means, though it clearly implies a picture-like display).

An argument along the following lines has been made in the recent neuropsychology literature (Kosslyn et al., 1999, see also the accompanying News of the Week article in *Science*, April 1999, Vol 284). Primary visual cortex (Area 17) is known to be organized retinotopically (at least in the monkey brain). So if the same early retinotopic visual area is activated when subjects generate visual images, it would tend to suggest that (1) the early visual system is involved in visual mental imagery, and (2) during imagery the cognitive system intercedes and provides the visual system with inputs in the form of a topographic display, like the one that is assumed to be normally provided by the eyes – in other words we generate a display that is laid out in a spatial or "depictive" form (i.e., like a two-dimensional picture). This interpretation was also supported by the finding (Kosslyn et al., 1995) that "smaller" images generated more activity in the posterior part of the medial occipital region and "larger" images generated more activity in the anterior parts of the region, a pattern that is similar to the activation produced by small and large retinal images, respectively.

There are plenty of both empirical and logical problems with this argument [5] which I will not address in this essay (but do address in Pylyshyn, submitted). For purposes of this essay, I will put aside these (often quite serious) concerns and assume that the conclusion reached by the authors of these recent studies are valid and that not only is the visual system involved in mental imagery, but also (1) a retinotopic picture-like display is generated on the surface of the visual cortex during imagery, and (2) it is *by means of this spatial display* that images are processed and patterns "perceived" in mental imagery. In other words I will assume that mental images literally correspond to *two-dimensional displays projected onto primary visual cortex to be reperceived by the visual system* in the course of reasoning about imaginary situations. We can then ask whether such a conclusion would help explain the large number of empirical findings concerning mental imagery (e.g., those described in Kosslyn, 1980) and thus help to clarify the nature of mental imagery. The purpose of this exercise is mainly to make the point that neuroscience evidence has no more claim to primacy in resolving disputes concerning mental processes than does behavior evidence, and indeed neuroscience evidence is of little help in clarifying conceptually ill-posed hypotheses, such as those being considered in the research on mental imagery.

What if we really found pictures in primary visual cortex?

Notice that what the neuropsychological evidence has been taken to support is the literal picture-in-the head story that people over the years have tried to avoid. It is no accident that the search for concrete biological evidence for the nature of mental imagery should have led us to this literal view. First of all, our search for neural evidence for the form of a representation can be no better than the psychological theory that motivates it. And the motivation all along has been the intuitive picture view. Even though many writers deny that the search is for a literal 2D display (e.g., Denis & Kosslyn, 1999), the questions being addressed in this research show that it is the literal view of images as 2-dimensional somatotopic displays that is driving this work. Secondly, if we were looking for support for a descriptivist view it is not clear what kind of neural evidence we would look for. We have no idea at all how codes for concepts or sentences in mentalese might be encoded. Even in concrete apparently well-understood systems like computers, searching the physical properties for signs of data structures would be hopeless. If our search was for a "functional space", which some people have suggested as the basis for images, we would still have no idea what to look for in the brain to confirm such an hypothesis. It is because one is searching for a literal 2D display that the research has focused on showing imagery-related activity in cortical Area 17 – because this area is known to be, at least in part, topographically mapped. The kind of story being pursued is clearly illustrated by the importance that has been attached to the finding described in (Tootell, Silverman, Switkes, & de Valois, 1982). In this study, macaques were trained to stare at the center of a target-like pattern consisting of flashing lights, while the monkeys were injected with radioactively tagged 2-deoxydextroglucose (2-DG), whose absorption is known to be related to metabolic activity. Then the doomed animal was sacrificed and a map of metabolic activity in its cortex was developed. This 2-DG map showed

[5] For example, there is serious doubt about the main premise of the argument, namely that primary visual cortex is essential for mental imagery, since there are clear dissociations between imagery and vision – even early vision – as shown by both clinical and neuroimaging data. And even if topographically organized areas of cortex were involved in imagery, the nature of the topographical layout of the visual cortex is not what we would need in order to explain such results as the effect of different images sizes on time to detect visual features (for example, larger images do not generate larger regions of activity, but only activity in different areas – areas that project from the periphery of the eyes – contrary to what would be required in order to explain the image-size or zoom effect, for example in the way it is explained in models such as that of Kosslyn, Pinker, Smith, & Shwartz, 1979).

an impressive retinotopic map of the pattern in V1, with only cortical magnification distorting the original pattern. In other words, it showed a *picture* in visual cortex of the pattern that the monkey had received on its retina, written in the ink of metabolic activity. This has led many people to believe that we now know that a picture in primary visual cortex appears during visual perception and constitutes the basis for vision. Although no such maps have been found for imagery, there can be no doubt that this is what the picture-theorists believe is there and is responsible for both the imagery experience and the empirical findings reported when mental images are being used. I have gone into these details because many people who cite the neuroscience results deny, when asked, that they believe in the literal picture view. But the lines of inference that are marshaled in the course discussing the evidence clearly belie this denial.

So we seem to be faced with the proposal, which is apparently supported by neurophysiological data, that when we entertain an image we construct a literal picture in our primary visual cortex which, in turn, is manipulated by our cognitive system and examined by our visual system. Given how widespread this view has become one ought to ask whether it makes sense on internal grounds and how well it fits the large body of data that has been accumulated over the past 30 years. What is the problem with this literal picture-view?

First of all, if images correspond directly to (or are isomorphic to) topographically-organized pictorial patterns of activity in the visual cortex, this pattern would have to be three-dimensional to account for the imagery data. After all, the content and function (as well as the phenomenology) of images is clearly three-dimensional; for example the same 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). Should we then expect to find three-dimensional displays in the visual cortex? The retinotopic organization of the visual cortex is not three-dimensional in the way required (e.g., to explain scanning and rotation in depth). 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. In fact, however, images are really multidimensional, insofar as they represent other spatially registered properties besides spatial patterns. For example, they represent color and luminance and motion. Are these also to be found displayed on the surface of the visual cortex?

Secondly, part of the argument for the view that a mental image consists of a topographical display in visual cortex is that the same kind of 2D cortical pattern plays a role in vision, so the visual system can play the dual role of examining the display in vision as well as in imagery. But it is more than a little dubious that visual processing involves examining such a 2D display of information about the visual world. It may well be that the visual cortex is organized retinotopically, but nothing follows from this about the form of the functional *mental* representations involved in vision. After all, we already knew that the retina started with a 2D display of activity, but nobody assumed that we could infer the nature of our cognitive representation of perceptual inputs from this fact. The inference from the physical structure of activity in the brain to the form of its functional representations is no more justified than would the parallel inference from a computer's physical structural to the form of its datastructures. From a functional perspective, the argument for the involvement of a picture-like structure in visual processing is at least as problematic as the argument that such a structure is involved in mental imagery. Moreover, the fact that our phenomenal percepts appear to be laid out in a phenomenal space is irrelevant because we do not see our internal representation, we see the world as represented and it is the world we see that appears to us to be laid out in space, and for a very good reason – because it is! We can easily be misled into believing that we are examining an internal display in vision just as we are in mental imagery, but both are illusions. The evidence is

quite clear that the assumption that an inner-display is constructed in vision is simply untenable (O'Regan, 1992; Pylyshyn, in preparation). Years of research on trans-saccadic integration have shown that our percepts are not built up by superimposing the information from individual glances onto a global image; indeed very little information is even retained from glance to glance and what is retained appears to be much more abstract and schematic than any picture (Irwin, 1996).

Thirdly, the idea that either vision or mental imagery involves examining a topographic display also fails to account for the fact that examining and manipulating mental images is qualitatively different from manipulating pictures in many significant ways. For example, it is the *conceptual* rather than *graphic* complexity of images that matters to how difficult an image superposition task is (see Palmer, 1977) and also to how quickly objects appear to be mentally "rotated", see Pylyshyn, 1979).[6] Although we appear to be able to reach for imagined objects there are significant differences between our motor interaction with mental images and our motor interaction with what we see (Goodale, Jacobson, & Keillor, 1994). Also accessing information from a mental image is very different from accessing information from a scene, as many people have pointed out. To take just one simple example, we can move our gaze as well as make covert attention movements relatively freely about a scene, but not on a mental image. Try writing down a 3 x 3 matrix of letters and read them in various orders. Now imagine the matrix and try doing the same with it. Unlike the 2D matrix, some orders (e.g., the diagonal from the bottom left to the top right cell) are extremely difficult to scan on the image. If one scans one's image the way it is alleged one does in the map-scanning experiment (Kosslyn, Ball, & Reiser, 1978), there is no reason why one should not be able to scan the matrix freely. Moreover, images do not have the signature properties of early vision; if we create images from geometrical descriptions we do not find such phenomena as spontaneous interpretation of certain 2D shapes as representing 3D objects, spontaneous reversals of bistable figures, amodal completion or subjective contours, visual illusions, as well as the incremental construction of visual interpretations and reinterpretations over time, as different aspects are noticed, and so on.[7]

I would turn this discussion of the parallels between vision and imagery around and suggest that the fact that in some situations the parallel between processing mental images and processing diagrams is so close it renders this entire line of evidence suspect, given that a real diagram and the way it is viewed using one's eyes has properties that no mental entity and process could have. Some of the psychophysical evidence that is cited in support of a parallel between vision and mental imagery entails a similarity that is so close that 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 if a real eye (Kosslyn, 1978) and that it has a field of resolution which is also the same as our eyes; it drops off with eccentricity according to the same function and inscribes the same elliptical resolution acuity profile as that of our (real) eyes (Finke & Kosslyn, 1980; Finke &

---

[6] It will not surprise the reader to hear that there are many ways of patching up a picture-theory to accommodate such findings. For example one can add assumptions about how images are tagged as having certain properties (perhaps including depth) and how they have to be incrementally refreshed from non-image information stored in memory, etc., thus providing a way to bring in conceptual complexity through the image generation function. With each of these accommodations one gives the actual image less and less explanatory work until eventually one reaches the point where the pictorial nature of the display becomes a mere shadow of the mechanism that does its work elsewhere, as when the behavior of an animated computer display is determined by an encoding of the principles that govern the animation..

[7] A great deal of research has been devoted to questions such as whether mental images can be ambiguous and whether we can make new construals of images constructed by combining other images. In my view the preponderance of evidence shows that the only reconstruals that are possible are not *visual* ones but inferences based on information about shape which could be in any form. These arguments are discussed in (Pylyshyn, submitted).

Kurtzman, 1981), and it exhibits the "oblique effect" wherein the discriminability of closely-spaced horizontal and vertical lines is superior to that of oblique lines (Kosslyn et al., 1999). Since in the case of the eye, such properties are due primarily to the structure of our retinas; these findings would suggest that the mind's eye is similarly structured! Does the mind's eye then have a blind spot as well? Of course, these close parallels could be just a coincidence, or it could be that the distribution of neurons and connections in the visual cortex comes 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* how things generally look 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 al the more plausible in view of the fact that the instructions in these imagery experiments explicitly ask observers to "imagine" that they are looking at a certain situation and to imagine what it would look like to see things, say, in their peripheral vision. The fact that subjects often profess ignorance of what would happen does not establish that they do not have tacit knowledge or simply memory of similar cases that they have encountered before (see note 2).

The picture that we are being presented, of a mind's eye gazing upon a display projected onto the visual cortex, is one that should arouse our suspicion. It comes uncomfortably close to the idea that properties of the external world, as well as of the process of vision (including the resolution pattern of the retina and the necessity of moving one's eyes around the display to foveate features of interest), are built into the imagery system. If such properties were built in, our imagery would not be as plastic and cognitively penetrable as it is. We can after all imagine almost any properties and dynamics we like, whether or not they are physically possible, so long as we know what the situation we are imagining would look like (we can't imagine a 4 dimensional world because we lack precisely this type of knowledge about it – we don't know where the contours, occlusions, shadows etc would fall). The picture-theory also does not even hint at a possible neural or information-processing basis for most of the interesting phenomena of mental imagery uncovered over the past several decades, such as the efficacy of visual mnemonics, the phenomena of mental rotation, and the apparent close parallels between how things work in the world and how we imagine them to work – which makes it possible for us to plan by visualizing a process and its outcomes. The properties exhibited by our imagery do not arise by magic: if we have false beliefs about how things work, our images will exhibit false dynamics. This is exactly what happens when we imagine light of different colors being mixed, or when we imagine an object in free fall. Because most people tacitly believe in the Aristotelian mechanics of constant-velocity free fall, our imagining of free fall is inaccurate and can be shown to follow the constant-velocity trajectory (for more such examples see Pylyshyn, 1981).

Where do we stand now?

Where, then, does the "imagery debate" stand at present? As I suggested at the beginning of this essay, it 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 the answer must be "yes"? If it is about whether in some sense imagery involves the visual system, the answer there too must be affirmative, since imagery involves experiences similar 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 questions are, of course; *what part* of the visual system is involved and in what way? Answering that will require a better psychological theory of the decomposition of the visual system itself. It is much too early and much too simplistic to claim that the way the visual system is deployed in visual imagery is by allowing it to *look at* a reconstructed retinotopic input of the sort that comes from the eye (or at least to some topographic remapping of this input).

Is the debate, as Kosslyn claims, about whether images are depictive rather than descriptive? That all depends on what you mean by "depictive". Is any representation of geometrical, spatial, metrical or visual properties depictive? If that makes it depictive then any description of how something looks, what shape and size it is, and so on, is thereby depictive. Does being depictive require that the representation be organized spatially? That depends on what restrictions are placed on "being organized spatially". Any physically instantiated representation is organized spatially – certainly both computer memories and books are. Does being depictive require that images "preserve metrical spatial information", as has been claimed (Kosslyn et al., 1978)? Again that depends on what it means to "preserve" metrical space. If it means that the image must *represent* metrical spatial information, then any form of representation will have to do that to the extent that spatial magnitudes need to be encoded and to the extent that people do encode them. But any system of numerals, as well any analogue medium, can represent magnitudes in a useful way. If the claim that images preserve metrical spatial information means that an image *uses spatial magnitudes to represent spatial magnitudes*, then this is a form of the literal picture theory. And a literal picture requires not only a visual system, but a literal mind's eye because the input is an uninterpreted layout of features.

Is there an intermediate position that we can adopt, somewhere between imagery being a symbolic representation and being a picture? This sort of representation has been the holy grail of many research programs, especially in artificial intelligence. In the case of mental imagery, the hope has been that one might develop a coherent proposal which says, in effect, that in mental imagery the visual system (or some early stage in the visual system) receives retintopically organized information that is nonetheless more abstract (or more conceptual) than a picture, but that still preserves a measure of spatial isomorphism. There is no principled reason why such a proposal could not work, if it could be properly fleshed out. But so far as I am aware nobody has even come close to making a concrete proposal for a type of representation (or a representational language) in which geometrical relations are encoded geometrically while other properties retain their symbolic force. Schemas, such as the mental models many people have discussed, represent special relations but do not have them. To have a geometrical relation would presumably require that the representation be laid out in some spatial medium, which gets us right back to the display view. The geometrical properties encoded in this way would then have to be cognitively impenetrable since they would be part of the fixed architecture. In any case this sort of "spatial schema" view of mental images would no longer be "depictive" in the straightforward intuitive sense. It would be more like a traditional semantic network or a schema, except that geometrical relations would be encoded in terms of spatial positions in some medium. Such a representation would have to be "read" just the way that sentences are read, except perhaps that proximity in the representation would have a geometrical interpretation (note that sentences too are typically encoded spatially, yet they do not use the space except to individuate and order the words). Moreover, such a spatial schema is unlikely to provide an account of such empirical phenomena as the ones described earlier – e.g., where smaller images take longer to see and distant places on an image take longer to scan to. But that is just as well since these are just the sorts of phenomena that are unlikely to be attributable to the nature of the image but to the knowledge that people have about the perceived world functions.

References

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375*(11), 121-123.

Denis, M., & Kosslyn, S. M. (1999). Scanning visual mental images: A window on the mind. *Cahiers de Psychologie Cognitive / Current psychology of Cognition, 18*.

Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *J Exp Psychol Hum Percept Perform, 6*(1), 126-39.

Finke, R. A., & Kurtzman, H. S. (1981). Mapping the visual field in mental imagery. *J Exp Psychol Gen, 110*(4), 501-17.

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., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3-71.

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.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5*(3), 94-100.

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. (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., Ganis, G., Sukel, K. E., & Alpert, N. M. (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*, 535-548.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual conrtex. *Nature, 378*(Nov 30), 496-498.

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.

Paivio, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Reinhart, and Winston.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9*, 441-474.

Pinker, S. (1980). Mental imagery and the third dimension. *Journal of Experimental Psychology: General, 109*(3), 354-371.

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. (1979). The Rate of 'Mental Rotation' of Images: A Test of a Holistic Analogue Hypothesis. *Memory and Cognition, 7*, 19-28.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*, 16-45.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

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. (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. (in preparation). *Seeing: It's not what you think*: Book ms.

Pylyshyn, Z. W. (in press). Visual indexes, preconceptual objects, and situated vision. *Cognition*.

Pylyshyn, Z. W. (submitted). Mental Imagery: In search of a theory. *Behavioral and Brain Sciences*.

Pylyshyn, Z. W., & Cohen, J. (1999, May, 1999.). *Imagined extrapolation of uniform motion is not continuous.* Paper presented at the Annual Conference of the Association for Research in Vision and Ophthalmology, Ft. Lauderdale, FL.

Shepard, R. N. (1975). Form, Formation, and Transformation of Internal Representations. In R. L. Solso (Ed.), *Information Processing in Cognition: The Loyola Symposium* . Hillsdale, N.J.: Erlbaum.

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.

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.

VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.

Acknowledgement: This research was supported in part by NIMH research grant 1R01-MH60924.