

title

If neuroimaging is the answer, what is the question?

authors

S. M. Kosslyn

institutions

832 William James Hall, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA (smk@wjh.harvard.edu)

It is unclear that we will come to a better understanding of mental processes simply by observing which neural loci are activated while subjects perform a task. Rather, I suggest here that it is better to come armed with a question that directs one to design tasks in ways that take advantage of the strengths of neuroimaging techniques (particularly positron emission tomography and functional magnetic resonance imaging). Here I develop a taxonomy of types of questions that can be easily addressed by such techniques. The first class of questions focuses on how information processing is implemented in the brain; these questions can be posed at a very coarse scale, focusing on the entire system that confers a particular ability, or at increasingly more specific scales, ultimately focusing on individual structures or processes. The second class of questions focuses on specifying when particular processes and structures are invoked; these questions focus on how one can use patterns of activation to infer that specific processes and structures were invoked, and on how processing changes in different circumstances. The use of neuroimaging to address these questions is illustrated with results from experiments on visual cognition, and caveats regarding the logic of inference in each case are noted. Finally, the necessary interplay between neuroimaging and behavioural studies is stressed.

keywords

Keywords: neuroimaging; visual imagery; visual perception; cognitive theory

heading

1. INTRODUCTION

Attending a poster session at a recent meeting, I was reminded of the old adage 'To the man who has only a hammer, the whole world looks like a nail'. In this case, however, instead of a hammer we had a magnetic resonance imaging (MRI) machine and instead of nails we had a study. Many of the studies summarized in the posters did not seem to be designed to answer questions about the functioning of the brain; neither did they seem to bear on specific questions about the roles of particular brain regions. Rather, they could best be described as 'exploratory'. People were asked to engage in some task while the activity in their brains was monitored, and this activity was then interpreted post hoc.

Sometimes, in the hands of a particularly talented investigator, this 'let's-try-it-and-see' approach can pay off. Usually, however, it does not; the result is a story made up after the fact, and many other stories could equally easily characterize the findings. In this article I wish to explore another approach to neuroimaging, one designed to answer specific questions. I wish to describe different types of questions that seem particularly well-suited to contemporary neuroimaging techniques, such as functional MRI (fMRI), positron emission tomography (PET) and magnetoencephalography (MEG). I shall illustrate approaches to most of the questions with examples from my laboratory, which is the research I happen to know best.

We shall consider two large classes of questions. The first class focuses on *how* information processing systems are organized in neural tissue. We ask which processes

Content

and structures confer specific functions, such as the ability to recognize objects or to form visual mental images. For present purposes, I use the word 'process' to refer to an operation or set of operations that transforms an input to an output, and 'structure' to refer to two types of entities: a 'representation' is a physical state that serves as a repository of information, and a 'buffer' stores representations. A metaphor might help: a blackboard is a buffer, marks made with chalk can serve as representations, and the movement of chalk on the board, the movement of an eraser on the board, and the act of interpreting the patterns of marks are all examples of processes. However, the distinction between structure and process is not clear-cut, as neural network modelling has demonstrated well: some structures actively transform input to output (see, for example, Kosslyn & Koenig 1995; McClelland & Rumelhart 1986; Churchland & Sejnowski 1992). Nevertheless, for present purposes this simple-minded characterization will serve as a useful organizational device.

The second class of questions focuses on *when*, on the specific circumstances under which, specific types of information processing are used. In some sense, one can answer such questions only if one has provisional answers to at least some questions of the first type. Once one has identified activation in a brain area with the operation of a structure or process, one can then ask in what circumstances such activation occurs. (I shall use the term 'activation' to refer both to increases in regional cerebral blood flow (rCBF) observed in PET and also to increases in signal strength observed in fMRI, which presumably reflect rCBF.) We look at two types of specific questions,

The first addresses whether one can infer the use of a particular structure or process on the basis of specific brain activation; the second examines whether one can use neuroimaging to consider ways in which context and other factors alter the use of a given structure or process.

Before launching into this far-reaching overview, it is useful to make explicit a key assumption: I am trying to organize types of question that one can sensibly try to answer using neuroimaging, but this does not guarantee that each such question will necessarily be answered by using these techniques. As we progress, I shall briefly note major methodological and conceptual stumbling blocks that can obstruct even the most clear-headed and systematic attack on an issue.

heading

2. HOW ARE INFORMATION PROCESSING SYSTEMS IMPLEMENTED IN NEURAL TISSUE?

Perhaps most of the current research with neuroimaging techniques is designed to reveal the functional architecture underlying a specific ability. By 'functional architecture' I mean the set of structures and processes that are used to perform a class of tasks. The areas activated while one uses the ability are taken to reflect the use of such structures and processes, and each area is characterized in terms of its role in implementing a particular structure or process. Sometimes researchers talk about a set of areas as a circuit, but this is usually misleading: in most studies all that is revealed are a set of activated (and/or deactivated) areas, with no information about the flow of information between the areas. Thus, what we see are the footprints of components of the functional architecture that are evoked during the task, but we do not see a specific circuit.

The initial questions that we will consider focus on a 'big picture' overview of a system of processes (setting aside for now details of how structures are involved with the processes), and later questions focus more narrowly on specific aspects of the system. I personally prefer to start with 'big picture' questions, which lead to research that provides a road map guiding the formulation of more specific questions.

(a) Which areas implement systems that confer specific functions?

Perhaps the most common type of neuroimaging study today is intended to reveal the neural events that confer a specific 'function' (i.e. ability), such as reading, object recognition or visualization. For any given function there are many different tasks that will engage some or all of the relevant structures and processes. I shall provide an example from our own work. We focused on the ability to identify objects, and examined the task of deciding whether a name is appropriate for a picture of an object. We began with a (coarse) theory of the functional architecture underlying such performance, and wanted to know whether it predicted the pattern of results. This theory is based in large part on research findings with non-human primates. Because I shall be using this work to illustrate my main points in what follows, I shall summarize the theory here (for a detailed treatment, see Kosslyn (1994)). As illustrated in figure 1, we posit a functional architecture with six major components, as follows.

The visual buffer is a set of topographically organized areas in the occipital lobe (see, for example, Felleman & Van Essen 1991; Van Essen 1985). These areas preserve, roughly, the spatial structure of images striking the back of the retina (and are therefore called retinotopically mapped areas), and have a key role in organizing visual input into perceptual units. The visual buffer serves both as a structure (as discussed below) and a process (by actively organizing input into perceptual units).

The attention window is a process that selects some localized portion of the visual buffer for more detailed processing (Cave & Kosslyn 1989; Posner & Petersen 1990; Treisman & Gelade 1980).

The pattern selected by the attention window is sent along two major pathways for further processing. The object-properties-encoding system analyses object properties (e.g. shape and colour) and runs from the occipital lobe to the inferior temporal lobe in the monkey; the human analogue seems to be the inferior temporal and/or middle temporal gyrus (see Haxby *et al.* 1991; Levine 1982; Mishkin *et al.* 1983; Sergent *et al.* 1992; Ungerleider & Mishkin 1982). There is good evidence that this system actually stores visual memories (reviewed by Kosslyn 1994). Input from the attention window is matched to these stored representations, leading to object recognition.

The second major pathway is the spatial-properties-encoding system, which encodes properties such as location, size and orientation, and runs from the occipital lobe to the posterior parietal lobe. There is considerable evidence that this system has a key role in guiding movements (see, for example, Goodale & Milner 1992). Not only are the areas in this system active when people or animals engage in spatial tasks, but also damage to it disrupts the encoding of spatial properties (see, for example, Andersen 1987; reviewed by Kosslyn 1994).

The two pathways must converge somewhere in memory; this inference is dictated by the simple fact that one can recall where things are located (e.g. the furniture in one's living room, objects on one's desk), which requires the cross-indexing of object and spatial properties. The object-properties-encoding system stores modality-specific visual representations. However, much of our knowledge is not modality specific; we can access this knowledge by multiple routes. For example, we can identify a cat when we see it, hear its meow, or feel it snuggling against our shin. This underlines an important distinction: *recognition* occurs when an input matches a stored unimodal representation, whereas *identification* occurs when we access information associated with that stimulus. When we recognize something, we know only that we have perceived it before, that it is familiar; when we identify something, we know that it has a certain name, belongs to specific categories, is found in certain locales, and so forth. Seeing a face and knowing that you have seen it before but not knowing when, where, or the person's name or other particulars, is an example of recognition without identification.

Such reasoning led us to posit an associative memory, which we hypothesize to rely on the angular gyrus and Area 19 (largely on the basis of clinical findings (see Kosslyn 1994)). Representations in associative memory are amodal ('abstract') and multimodal, and store

associations to objects, scenes and events. These representations are activated by converging input from the object-properties-encoding and spatial-properties-encoding systems.

In situations in which an object cannot be recognized well (because the input does not match well any previously stored representations), the best-matching representation in associative memory is treated as a hypothesis about the identity of the stimulus. The information look-up process accesses stored information associated with the hypothesized object, and uses such information to guide a top-down search. The information look-up process essentially shunts information from one part of the brain to another.

According to our theory, information stored in associative memory is used during a top-down search not only to guide the mechanisms that shift attention (to look for a distinctive part or characteristic that should be present if the hypothesis about the stimulus' identity is correct), but also to 'prime' representations in the object-properties-encoding system. One of the more striking facts about the neuroanatomy of the visual system is that virtually all areas connected by afferent fibres are also connected by efferent fibres (i.e. the connections are reciprocal); there is an enormous amount of information flowing backwards in the visual system (see, for example, Felleman & Van Essen 1991). We conjecture that the efferent connections are used for 'cooperative computation', and that the sort of priming posited here is at the heart of such processing. This sort of priming facilitates the encoding of an expected part or property (possibly by lowering thresholds, or by increasing sampling rates). Dorsolateral prefrontal cortex is critical in implementing the information look-up system (see, for example, Damasio 1985; Goldman-Rakic 1987, 1988; Luria 1980; see also Kosslyn *et al.* 1995b).

Finally, there are processes that actually shift attention. These *attention shift* processes are implemented in several parts of the brain. Posner & Petersen (1990) distinguish between three different components of this system, one in the posterior parietal lobe that disengages attention from its current location; one in the superior colliculus, frontal eye fields and other neural structures that guides the attention window, as well as the eyes, head and body to a new location; and one in the thalamus that re-engages attention at a new location.

Kosslyn *et al.* (1994) set out to discover whether this theory could be generalized to humans. In our view, a good neuroimaging experiment hinges on comparing two tasks that are minimally different. In most cases, it is best if only a single aspect of the task is varied at a time; if more than one characteristic is varied, the variables should be manipulated orthogonally (keeping everything else constant while manipulating each one). Thus, in this study we examined the same task, deciding whether words name pictures, but in two conditions: in one, the pictures were presented from canonical (typical) points of view, whereas in the other the pictures were presented from non-canonical (atypical) points of view. Counterbalancing over subjects ensured that the same pictures and words were used equally often in both conditions. Our reasoning was that the canonical views would match stored representations in the object-properties-encoding

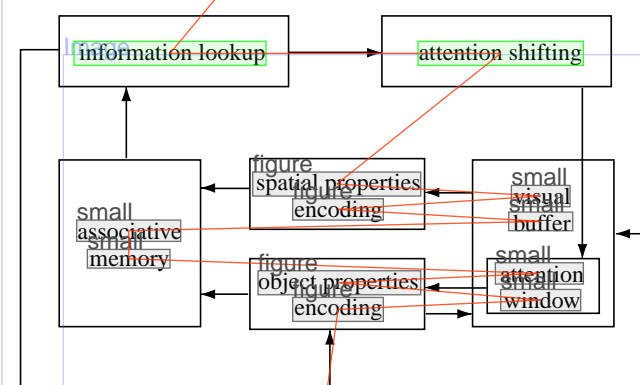
system, and hence be recognized well, and then would be identified immediately. In this case, although all of the other processes would be running at the outset, they would not need to run to completion to perform the task. (Figure 1 is misleading in that it might suggest that each process waits for the products of the previous one before operating. On the contrary, there is good reason to assume that all parts of the brain are 'running' at the same time, and that what changes during task performance is how intensively each one is operating. When processing shifts over time, this is reflected by increased activity in areas that implement the relevant processes.)

In contrast, when faced with a non-canonical view, the object would not be recognized with confidence initially (if it were novel, as ours were). Thus, it would be identified only tentatively, and this best-matching representation in associative memory would serve as the basis for a hypothesis. At this point, the information look-up process would continue to operate, looking up characteristics associated with this hypothesized object, and would pass information needed to shift attention appropriately and to prime the object-properties-encoding system to encode the expected characteristic. A characteristic at the specified location would then be encoded into the visual buffer, and would thereafter be processed in the object-properties-encoding and spatial-properties-encoding systems, which would then provide input to associative memory. If the expected part or characteristic was in fact located (as indicated by a match in the object-properties-encoding system) and was located in the correct place (as registered in the spatial-properties-encoding system), this might be enough input to associative memory to identify the object. If not, an additional cycle might be necessary.

In short, we expected the entire system to cycle more times for the objects seen from non-canonical perspectives; and in fact, we found activation in areas that correspond directly to those subsumed by figure 1. The only mystery was activation in what we initially took to be the motor strip; however, subsequent evidence (see, for example, Paus 1996), has shown that this area probably corresponds to the human frontal eye fields.

However, many areas were activated, and one cannot help wondering what sort of results would have disproved the theory. One can make two (non-exclusive) moves at this point. First, a set of results such as ours can be viewed as a set of structured hypotheses. One has now identified areas that are theorized to perform specific roles in information processing (such as matching visual input to stored modality-specific visual memories, or encoding spatial properties), and hence one now knows where to look for additional, more focused evidence (as I shall discuss in the next two sections).

Second, one need not study the putative function of each area separately, but rather one can design studies of the entire system that rely on different tasks. Ideally, the tasks should on the surface not appear similar, and thus one can have confidence that the interpretation of the functions of the areas is coherent if the same areas are activated in both tasks. For example, Kosslyn *et al.* (1997) expected the system of figure 1 also to be used when one forms visual mental images. In this case, one 'primes' the object-properties-encoding system so strongly that



imageDescription
Figure 1. Major structures and processes used in later visual perception and visual mental imagery, according to the theory of Kosslyn (1994).

Content
efferent connections literally force a pattern of activation in the areas that implement the visual buffer. This pattern of activation corresponds to the mental image, which once present can in turn be scanned by the attention window, portions recognized individually, locations noted, and so forth, in the same way that occurs during visual perception. For example, when asked to report the number of windows in their living room from memory, most people claim that they visualize the room and shift their attention along the walls, 'seeing' each window in turn. In the course of counting, they may notice characteristics of the windows (such as the types of locks that are present) that they had not previously considered. The information was implicit in the representations in visual memory (in the object-properties-encoding system) but never identified. Similarly, if asked what shape are a cat's ears, most people visualize the animal's head and 'look' at the ears, categorizing their shape. If asked this question several times in succession, they memorize the answer explicitly and no longer need to use the image (see Kosslyn 1980).

In addition, if multiple parts need to be added to an image, one can visualize an overall shape envelope, and then shift attention to the location of a part to be added, just as one would during top-down hypothesis testing during perception; however, in this case, once one is focused on the appropriate location, one forms an image of the part (via the priming mechanism, as just described).

Kosslyn *et al.* (1988) used a task originally designed by Podgorny & Shepard (1978) to test the idea that visual mental images are built up a part at a time. In this task, subjects see a 4×5 grid that contains a single X mark in one cell. The subjects are asked to visualize a block letter in the grid, and report whether the X mark would have fallen on part of the letter if it were actually present. Kosslyn *et al.* (1988) found that the time to respond depended on the location of the X mark. In fact, when other subjects were surreptitiously observed drawing block letters in such grids, they typically drew the segments in very consistent orders. For example, for an upper case letter F, they drew the vertical line on the left first, the top horizontal line second, and the middle horizontal line third. The order in which these subjects drew the segments neatly predicted how quickly other

subjects could determine whether an X mark would have fallen on the letter if it had been in the grid. Indeed, the more segments one would need to draw to reach the location of the X, the longer it took to determine whether the X would have fallen on the letter if it had been in the grid (see also Kosslyn 1988). This result is exactly as expected if one were in fact sequentially shifting attention, and forming an image of each segment in sequence.

The theory outlined in figure 1, then, led us to expect the same set of areas to be involved in the two tasks, identifying named pictures seen from non-canonical versus canonical points of view and deciding whether an X mark would fall on an imaginary letter. Thus, Kosslyn *et al.* (1997) replicated the experiment of Kosslyn *et al.* (1994) and also asked the subjects to perform the image-in-grid task. The control condition for the imagery task was the identical set of stimuli, but presented before subjects had seen the block letters or knew anything about the imagery task. For the control, subjects simply responded as soon as they saw each stimulus, alternating the side of response over trials. The results revealed that two thirds of the areas activated during either task were activated in common. The functions of these areas are organized well by the theory outlined in figure 1.

However, we must note that to address this type of question, one must design pairs of tasks that differ in a single way, which is not as straightforward as it might seem. One problem is that we cannot know in advance whether manipulating a single feature of the task in fact alters only one aspect of processing. As noted by Kosslyn *et al.* (1994), it has long been known that changing one aspect of a task can lead subjects to adopt qualitatively different strategies. Nevertheless, in my view this approach is still superior to using rest (or an 'off' state) as the baseline; a rest condition surely differs in more than one way from most experimental conditions, and it is impossible to know in which ways the conditions differ.

In any event, the results of this sort of research, in my view, can only be viewed as preliminary. They provide some support for a structured set of hypotheses, but to be compelling each aspect of the overall system must be studied in more depth individually. However, the virtue is that these more precise and circumscribed hypotheses are not conceived in a vacuum, isolated from conceptions of the operation of the system as a whole. The theory outlined above treats each structure and process as part of an integrated system, and this seems highly desirable for questions about how the brain performs information processing.

heading (b) Which areas implement 'simple systems' (linked operations)?

A second kind of question focuses not on the operation of an entire system that underlies a particular function, but instead on a portion of such a system. The simplest subsystem is the interaction between two processes or between a process and a structure. For example, perhaps the most controversial aspect of the theory illustrated in figure 1 is the interpretation of the dorsolateral frontal function as being involved in information look-up, and of the angular gyrus or Area 19 areas as implementing associative memory. Kosslyn *et al.* (1995b) tested this

interpretation. We made use of a finding originally reported by Jolicoeur *et al.* (1984).

Jolicoeur *et al.* (1984) asked subjects to view pictures and to decide whether simultaneously presented words described the pictures appropriately. There were two important variables: first, the words were either at the 'entry' level (such as 'dog' for a picture of a cocker spaniel, or 'apple' for a picture of a Delicious apple), or were superordinate to that (e.g. 'animal' or 'fruit') or were subordinate (e.g. 'cocker spaniel' or 'Delicious apple'). The idea was that people spontaneously name objects at the 'entry' level, which usually corresponds to the 'basic level' of Rosch *et al.* (1976); this is the name that is as general as possible and that names as many similar objects as possible. So, for example, 'Delicious apple' is too specific because 'MacIntosh apple,' 'Granny Smith apple,' and so on look similar, but 'fruit' is too general because watermelons, grapes, bananas and so on are not very similar; 'apple' is about as general as one can get while still naming a set of similar objects. As expected, subjects could evaluate an entry-level term faster and more accurately than either a subordinate or superordinate term (see Smith & Medin 1981).

Second, the pictures were presented for either 1s or 75 ms. We had two pivotal ideas. (i) If one names a picture spontaneously at the entry level, then one will need to process additional perceptual information to evaluate a subordinate term (e.g. to check whether there are bumps on the bottom of an apple to determine whether it is a Delicious apple, or to see whether the dog has rounded floppy ears). If so, then it should be much more difficult to evaluate subordinate terms than entry-level ones when given only 75 ms than when given a full 1s. This was in fact so. (ii) In contrast, and here is the crucial idea for present purposes, if one names a picture spontaneously at the entry level, then one will not need to process additional perceptual information to evaluate a superordinate term. In this case, the additional information need only be looked up in associative memory. For example, once one has named a Delicious apple as 'apple,' one needs only to search memory to discover that apples are indeed members of the category 'fruit'. If so, we reasoned, then the exposure time of the pictures should have comparable effects on one's ability to evaluate entry and superordinate terms, because one did not need to encode additional perceptual information after producing the entry-level name. In fact, this was true. Although subjects required more time to evaluate superordinate terms, they required the same amount of additional time compared with entry-level terms in the two exposure conditions.

Given these findings, Kosslyn *et al.* (1995b) reasoned that if we used PET to compare the additional activation when subjects evaluated superordinate terms compared with when they evaluated entry-level terms, we should find just the areas involved in looking up information from associative memory. Indeed, Kosslyn *et al.* (1995b) found only two areas to be more active during the superordinate condition than during the entry-level condition, namely the left dorsolateral prefrontal cortex and the left angular gyrus. We expected only left-hemisphere activation because the subjects were evaluating words, and therefore the results make good sense in the context of the theory.

These findings were very welcome, but they were not ideal. The specific loci that we found were not precisely the same as those found in the original experiment. However, these were different subjects, and different averaged data. Nevertheless, the locus of the dorsolateral prefrontal area was considerably different in this and the studies described above, which led us to expect activation consequent upon use of the information look-up process and associative memory. It simply is not clear what to make of this; the *a priori* predictions were clear-cut, and the results were close to the expected areas; but how close is close enough? A major advantage of fMRI is that it is possible to perform many experiments with the same subjects, and to analyse each subject's data separately. If the variability that we noted between the different PET studies was simply sampling error, then for a given person precisely the same dorsolateral prefrontal cortex and posterior areas should have been activated in the original tasks (both picture-verification and imagery) and in this comparison between superordinate and entry-level terms.

It is important to recognize two important assumptions that underlie the logic of this sort of research. First, as noted above, changing the task in small ways can alter the strategy that subjects use, in which case different structures and processes might be used, and hence different areas might be activated. Even a small shift in the location of an activated area could reflect differences in the precise process being used (see, for example, Wilson *et al.* 1993). Second, note that this logic assumes that the same piece of tissue implements the same structure or process when it is embedded in different combinations of other processes. It is possible that some neural tissue is not dedicated to implementing a single process, and that different neural regions can implement more than one process. If so, the precise locus of activation that reflects the use of a given structure or process can shift, depending on what other structures or processes are used at the same time (for a more detailed discussion of this point, see the epilogue of Kosslyn & Koenig (1995)). We shall return to this general issue shortly.

(c) What operation is performed by a specific brain area?

One also can attempt to characterize the process or processes implemented in a single area. It is clear that any task at all, even one as simple as staring at a fixation point (Petit *et al.* 1995), involves a neural circuit that draws on multiple brain areas. Moreover, given our uncertainty about the precise relationship between manipulating a variable in a task and changes in underlying processing, it is difficult to design true 'minimal difference' pairs of tasks, so that only a single process is taxed more by the experimental task than the control task. Instead, the most sensible course for answering such questions has been plotted by Price & Friston (1997), in their 'conjunction analysis' investigations. In these studies they administer multiple tasks (but as few as two) and examine areas of common activation (i.e. that are activated strongly enough to generalize across tasks in a location). For example, they have used this approach to study the shared processing when one names pictures, letters and colours, and reads words. Although their work so far has revealed not individual areas, but sets of areas,

the same approach can in principle allow one to characterize in detail the function of individual areas.

However, one problem with all such research rests on the logic of 'affirming the null hypothesis'. One implicitly is assuming that a large host of untested stimuli would have no effect. That is, if one shows that area X responds when the subject is engaged in process Y, this does not mean that area X would not also respond when the subject was engaged in the as-yet-untested process Z. Thus, one's interpretation of what the area does depends on the questions that one puts to it; given only certain stimuli in certain tasks, one can be misled (e.g. into believing that the area has a more specialized function than in fact it does). As far as I can see, there is no way around this problem other than continuing to refine one's theories and one's tests, eventually eliminating possibilities and narrowing down the range of viable interpretations of the function of the area in question. The best of all worlds, it seems to me, is to have a theory that makes a very specific prediction and to consider the alternative theories in advance, ensuring that they do not also make the same prediction.

heading (d) Which are the properties of structures that underlie a particular ability?

For present purposes, I restrict the term 'structure' to refer to a buffer. A buffer is a physical medium that can retain representations. A representation, in the simplest terms, is a physical pattern that stands for something else. For a physical pattern to count as a representation, it must occur within a system of processes that operate on the pattern. For example, the 'meaning' of patterns of magnetic flux in a computer's random-access memory (RAM) is determined by the way in which the patterns are 'read' by the central processing unit and the impact that such processing has on the production and modification of yet other patterns of magnetic flux. In this case, the RAM serves as a buffer, which supports physical patterns that represent data. (Note that a program, in this sense, is just data: it is a specification of a sequence that the central processing unit should perform.)

Neuroimaging can be used to ask questions about the nature of structures. For example, Fox *et al.* (1986) used PET to study the organization of Area 17 (primary visual cortex). In particular, they wanted to collect evidence that human Area 17 is topographically organized. In topographically organized areas, nearby points of space are represented by nearby neurons, and more distant points of space are represented by more distant neurons. (This is true within a given 'cortical magnification' factor; the scaling of input to cortex becomes more coarse towards the periphery, and this must be taken into account when interpreting the physical pattern.) Topographically organized areas are 'picture-like,' but typically are not perfectly isomorphic maps; there are various distortions, such as greater tissue being devoted to the input from the high-resolution foveal parts of the eye. Fox *et al.* asked subjects to view alternating chessboards that just fell on the fovea, were larger but did not stimulate the fovea (these were doughnut-shaped patterns, with a hole in the middle), or were larger still and did not stimulate an even larger central area. They found that the larger the stimulus, the more anterior was the focus of activation

in Area 17 (i.e. along the calcarine sulcus). This pattern of responses was just as expected if this area were retinotopically mapped, as it is in animals. However, unlike some animals (such as the macaque monkey, which has Area 17 on the lateral surface), in humans this area is medial and the cortex is folded in complex ways. Thus, the PET results were interesting because they revealed the expected functional correlates of what was known about the physical structure of the area (and what had been suggested by earlier studies of visual deficits after local damage to this area; reviewed by Kosslyn 1994).

The study by Fox *et al.* (1986) suggested to me to ask the corresponding question about visual mental imagery. There has been a long debate about whether a picture-like representation underlies the experience of visual mental imagery (see, for example, Kosslyn 1980, 1994; Pylyshyn 1981; Tye 1991). Neuroimaging can be used to ask whether such a structure is evoked during visual imagery, as suggested in figure 1. For example, Kosslyn *et al.* (1995a) asked subjects to visualize line drawings of common objects under three different conditions. The conditions were the same, except that a different-sized box was shown to the subjects before each, and the subjects were asked to visualize each named drawing during that condition as if it were within the box. Thus, they visualized objects at a very small size (0.25° of visual angle, from their perspective), at a medium size (4°) or at a very large size (16°). After visualizing a drawing, subjects were asked to evaluate a subtle shape property (such as whether in that drawing the left side of the object had been higher than the right), which required them to use imagery (they did not know about the particular judgements to be asked at the time at which they studied the drawings). During scanning, the subjects had their eyes closed and all cues (names of objects to be visualized and shape probes) were delivered auditorily. In the 'listening baseline,' they heard the same kinds of stimuli used in the imagery conditions, but now simply responded when they heard a shape probe; this control was always conducted before the imagery conditions, so the subjects did not know the significance of the terms. In fact, there were four sets of recorded object-probe stimuli, which were rotated so that each occurred equally often in each of the three imagery conditions and in the listening baseline. The data from the imagery conditions were also compared with those from a 'resting' baseline condition, but this is not relevant for present purposes.

The results were straightforward. When the imagery data were compared with those from the listening baseline, we found activation in Area 17. Moreover, the locus of maximal activation depended on the size at which the drawings were visualized. We found very posterior activation with the small images, more anterior activation for the medium-sized images, and still more anterior activation for the large images. These results parallel those found in studies of visual perception, and are as expected if this topographically organized structure (part of the visual buffer illustrated in figure 1) is in fact used in visual imagery.

It is important to note that the simple finding that an area is activated, even an area with well-characterized properties, is not sufficient to infer that the properties of that area contribute to performance. As noted earlier, the

choice of a baseline is crucial, and it is possible that a test condition can seem to produce activation because the baseline condition actually reduces activation (as might happen in visual cortex when one is concentrating on listening to cues). Thus, in the study just described it is important that the imagery results cannot be ascribed to hypometabolism during the baseline condition: this would not explain the selective effects of size in itself on the locus of activation. By varying the sizes of objects in images, we could show directly that the spatial properties of this structure are drawn upon during performance of the task. In addition, we have found that if repetitive transcranial magnetic stimulation (TMS) is applied over Area 17, which has the effect of temporarily impairing the underlying tissue, subsequent imagery performance is impaired (Kosslyn *et al.* 1999); thus, in this kind of task, the activation in early visual cortex has a functional role. In short, this research reveals that a spatially organized buffer can in fact be used during visual mental imagery.

In addition, this use of neuroimaging can reveal distinctions where none were previously apparent. In this example, other research results have failed to reveal the activation of Area 17 during visual imagery (for recent reviews, see Mellet *et al.* (1998) and Thompson & Kosslyn (1999)). Many laboratories have now found one or the other result, and thus neither the presence nor absence of the effect can be explained away. Rather, it seems clear that there is more than one type of imagery, and not all types rely on the kind of depictive representation that is associated with our introspections of imagery. This finding does not detract from the fact that a topographically organized structure is used in some types of imagery.

Headline

(e) *Is there more than one way in which a function can be performed?*

In the previous sections we considered how one can discover the processes and structures that give rise to a specific ability. We initially considered studies that provide a 'big picture' overview of the system that underlies a particular ability, and then narrowed our focus to increasingly small portions of such a system. One virtue of starting in this way is that it puts the more precise questions in context; it leads one to ask questions about how processing occurs in a system. Each process depends on others for its inputs, and in turn sends its outputs to yet other processes (and structures). Thus, to begin to think about what an individual process or structure does, it is useful to think of it in context.

Armed with such information, it is straightforward to use neuroimaging to consider another type of question, that of whether there is more than one way to perform a particular function. That is, can different combinations of structures and processes underlie a particular type of performance? Each such combination of structures and processes corresponds to a different strategy. The key is to show that different strategies produce different patterns of activation. For example, Kosslyn *et al.* (1998) asked whether there is more than one way in which to rotate objects mentally in mental images, one that draws on motor processes (as would occur if one visualized what one would see if one twisted an object in a certain way) and one that does not (as would occur if one visualized what one would see if an external force twisted an object).

In this study, subjects performed two types of mental rotation tasks. In one, the subjects performed the classic Shepard & Metzler (1971) task. They saw pairs of three-armed angular shapes, half of which were identical and half of which were mirror images. We administered two conditions. In the test condition, the shapes appeared at different relative orientations, and subjects decided whether the members of each pair were identical or mirror images. We compared brain activation in this condition to that in a baseline condition, which was identical to the test condition except that the shapes in each pair were always presented at the same orientation. In the second mental rotation task, the subjects saw pictures of pairs of hands, and had to decide whether both were the same hand (right or left) or one was right and one was left. In the test condition the hands appeared at different relative orientations, whereas in the baseline condition both members of the pair appeared at the same orientation. Each subject received the baseline and test conditions for one type of stimulus before the two conditions for the other type, and the order of presentation was counterbalanced.

Kosslyn *et al.* (1998) reported that areas involved in motor processes were activated in the hands condition, but not with the Shepard & Metzler stimuli. This finding provided support for the idea that there are at least two distinct ways in which to imagine objects rotating, one that involves motor processes and one that does not. These findings are all the more compelling because prior data showed that some of the areas that were activated during rotation of hands are in fact used in motor control, and showed that other areas were not so involved. Thus, although the mere fact that there were distinct patterns of activation in the two tasks is evidence that they are performed in different ways, the ability to interpret the differences adds meat to these bones. A plausible interpretation allows one to eliminate the possibility that the differences were due to chance or due to some incidental difference in the experimental designs.

A crucial aspect of this logic is that the patterns of activation arising from the two strategies must actually be different. It is not enough to show that, relative to a baseline, in one task areas X–Y are activated, whereas in another task areas P–Q are activated. 'Activation' is a continuous variable, and sub-threshold activation is common. Thus, even though activation in an area might be significant in one task (relative to baseline), and not in another, there might be no difference between the two when they are directly compared. Thus, one must compare the tasks directly, not simply rely on visual comparisons of what was and was not over the threshold under separate conditions.

heading

3. WHEN ARE SPECIFIC TYPES OF INFORMATION PROCESSING USED?

One virtue of beginning by asking questions about how information processing is implemented in the brain is that the answers can later be used as a basis for asking and answering other types of questions. We can distinguish two broad classes of questions that can be asked about when specific processing is used. First, one can ask simply whether a given process or structure was drawn upon

while subjects engaged a particular ability. Second, 'when' need not be a constant: one can ask how practice, fatigue or some other state can alter when a particular process or structure is used. We explore both classes of questions below.

heading

(a) *Inferring processes from brain activation*

Two distinct types of logic of inference can be used to infer that a particular process or structure is used when subjects perform a specific task.

heading

(i) *Inferring processing from the presence of activation*

It is commonplace in the discussion sections of neuroimaging studies to see post-hoc interpretations of processing based on the set of areas found to be activated. In the ideal case, if one knows which processes or structures are implemented in a given part of the brain, one can argue subsequently that activation in that area while subjects perform a particular task is *prima facie* evidence that these processes or structures were drawn upon while subjects performed that task. To the extent that a particular process or structure has been identified with a particular neuroanatomical locus, this logic of inference is defensible. For example, the function of area M1 is understood well enough to allow us to infer that if it is activated, then motor processes were present as one performed the task (assuming that it can be shown that the activation is not an artefact of the baseline comparison).

For example, it is known, largely from animal research (see, for example, Checetto & Saper 1990), that the anterior insula is the major cortical recipient of input from the autonomic nervous system. We therefore asked whether visual mental images of strongly aversive pictures could induce such feedback from the body; it is well-known that mental images can engage the autonomic nervous system (see, for example, Lang *et al.* 1980). Kosslyn *et al.* (1996a) asked subjects to visualize either very aversive pictures (e.g. of a baby with a large tumour over one eye, or a battered body) or neutral pictures (e.g. of a truck, a lamp), mostly drawn from Lang's set (Lang *et al.* 1988). PET scanning was conducted while subjects visualized each picture and determined whether a statement was an appropriate description of it; the statements were selected to require imagery to evaluate, and thus to encourage subjects to visualize the pictures (e.g. 'the baby has a tumour over his left eye'). When rCBF in the aversive condition was compared with that in the neutral condition, additional activation was documented bilaterally in the anterior insula, as expected. (Other areas, such as Area 17, were also activated, but this is not to the present point.) This led us to infer that images of aversive objects can engage the bodily structures that in turn feed back to activate this area.

We can discover whether the structures or processes implemented in a particular area have a functional role in conferring an ability by studying the effects of disrupting that area. In this case, patients who had damage to areas that have no essential role in the ability of interest would still retain the ability (assuming that they are alert enough to be engaged in the tasks used to assess the ability); similarly, TMS delivered to the area would not disrupt this ability. In contrast, if an area does indeed underlie a particular ability, the ability should be

impaired after both damage to the area and temporary disruption of the area induced by TMS (Pascual-Leone *et al.* 1997; Kosslyn *et al.* 1999). However, without such data, there is no way to know whether any given area implements a structure of process that contributes to the performance of the task of interest.

In addition, a problem with the logic of inferring processing from activation is that a given part of the brain can implement more than one process. This could conceivably be true even for highly constrained sensory and motor cortices; we simply do not know enough to exclude the possibility of multiple roles for any piece of cortical real estate. Thus, the activation of a given brain area does not necessarily implicate specific processes that were previously found to activate it; some other process could in fact be at work. The finding of activation in a particular brain area during a particular task can therefore only be taken as either (i) the basis for the hypothesis that a specific process or structure is at work (to be tested, in part, by observing the effects of disrupting the area), or (ii) one piece of converging evidence for such an inference.

heading

(ii) *Using variations in activation to predict performance*

A completely different logic of inference can be used to discover whether a specific process or structure contributes to how well one can make use of a specific ability. First, however, it is important to be clear on what question is being asked here. We are no longer asking whether a process or structure (implemented in an area) confers an ability. Instead, we are asking whether variations in the efficacy of a structure or process are related to variations in the performance of the relevant tasks. That is, we can divide processes and structures into two general types. The first type of processes or structures are not taxed by the task, to the extent that, if they function at all, they function well enough to allow one to perform a class of tasks. The second type are the 'rate-limiting steps'. They are taxed when one draws on a particular ability; the more effectively they function, the better one will perform the relevant tasks. By analogy, keys on a keyboard are designed so that minimal strength is required to press them down; therefore, finger strength in excess of this amount is irrelevant (and hence finger strength would be analogous to the first type of process). However, the ease of typing a sequence of letters depends on their arrangement, and setting up motor commands to execute some sequences is difficult; thus, better ability to programme complex motor movements will lead to better performance on this task (this is a rate-limiting step).

For example, in the task of Kosslyn *et al.* (1996b), subjects were asked to visualize, with eyes closed, capital letters. After 4 s they heard a probe (such as 'straight side') and were to decide, as quickly and accurately as possible, whether the visualized letter had that property. Both response times and accuracy rates were recorded as the subjects performed the task while being PET-scanned. After both behavioural and rCBF data had been collected from 16 subjects, the rCBF values in each brain were normalized to the same mean. For each brain, we then measured deviations from the mean for each region of interest (ROI) found previously to be activated during visual imagery (as summarized in §2). These areas were previously interpreted as implementing the functions

outlined in figure 1. We next performed stepwise multiple regression analyses, using either response times or error rates as the dependent measure and ROI values as the independent measures. The error rates were very low and varied only slightly between subjects, so no significant correlations emerged, but the response times were of great interest. The rCBF values of three areas were entered into this equation, each accounting for significant amounts of variance in the response times. The first area was Area 17, which by itself was correlated at $r = -0.65$ with response times. The other two areas were Area 19 (which we take to implement part of associative memory) and the posterior parietal lobe (which we take to be part of the spatial-properties-encoding system). The multiple R value was 0.93, which is remarkably high given how different measures of rCBF are from response times.

Thus, the fact that variations in rCBF in those three areas contributed to variations in performance is evidence that those areas have a role in performing the task. (For comparable results regarding the role of the amygdala in encoding emotional memories see Cahill *et al.* (1996), and for comparable results regarding the role of the hippocampus and related structures in episodic memory see Nyberg *et al.* (1996).)

One unexpected finding in Kosslyn *et al.* (1996b) was that the correlation between parietal rCBF and response time was positive: the more blood flow there was, the more time was needed to respond; in contrast, the other two correlations were negative (the more blood flow, the less time needed to respond). This finding might suggest that people who performed the task relatively poorly might be using a different, less effective, strategy than those who performed it relatively well. Those who performed well seemed to rely on Area 17, which supports visual images of shapes, whereas those who performed poorly had relatively little rCBF in this area. Instead of relying on visual images, the subjects who performed poorly might have relied on spatial representations that are used to guide drawing the segments, and these representations were accessed via processes implemented in the parietal lobes.

A major virtue of this approach is that no subtractions or comparisons between tasks are necessary. One can design tasks so that there is a single 'rate-limiting' step, and simply correlate responses with rCBF values. However, one must have some prior reason for looking at particular ROIs. In addition, this approach will not allow one to identify all the structures and processes involved in performing a task. Furthermore, as our results should make apparent, it is not entirely clear how to interpret the direction of the correlation between rCBF and a measure of performance. Moreover, if people differ widely in their skill levels, the correlations might be even more difficult to interpret: low values of rCBF for some people might indicate that the process is not performed well, but low values for other people might indicate that they are so expert that only minimal processing is required. Again, results from this method must be viewed as only one more source of converging evidence.

heading

(b) Changes in processing

We have been considering questions to address when a particular process or structure is used. However, 'when'

need not be a constant, but can change depending on a variety of factors. Three such factors in particular seem worth consideration.

heading

(i) How does processing change with practice?

Results from several laboratories have shown that practice does not simply lower rCBF values in areas used when one is performing a given type of task. Rather, entirely new areas can begin to function as one becomes better at performing a task, and areas activated initially can decrease activity below the threshold of detection. For example, we recently performed a simple study (M. T. Menard, W. L. Thompson, P. I. Bentley, L. L. LeSueur, N. M. Alpert, S. M. Kosslyn and A. M. Galaburda, unpublished data) in which subjects were asked to read words aloud. They did this four times, with the particular words being used in each condition being counter-balanced over subjects. As expected, many areas are activated when one reads for the first time compared with the fourth time, but the reverse was also true: additional areas became activated by the fourth blocks of trials that were not activated initially (for comparable results, see Petersen *et al.* (1998) and Raichle *et al.* (1994)).

In addition to investigating how processing changes with practice, one can ask how this change varies for different populations. For example, the reading-aloud study (M. T. Menard, W. L. Thompson, P. I. Bentley, L. L. LeSueur, N. M. Alpert, S. M. Kosslyn and A. M. Galaburda, unpublished data) compared practice in reading in normal populations and in dyslexics (who were age- and education-matched to the normal control subjects). We found again that some areas activated initially dropped out with practice and that different areas became evident. However, the particular areas were strikingly different from those noted in the controls, in both cases: initially and after practice.

Findings such as these are interesting only if they can be interpreted. The interpretation relies on answering questions of the sort that we considered in the first major section of this article. However, this need not be a one-way street. Observations about changes of the sort noted here can lead to hypotheses about what process or structure is implemented in specific areas, and these hypotheses can in turn be examined in ways described above. This 'golden feedback loop' seems more likely to occur if one comes armed with a set of questions from the outset, based on a theory, than if one simply performs neuroimaging while subjects perform a task and then interprets the results post hoc.

heading

(ii) How does processing change with context?

We also can ask how processing changes in different contexts. Let us return to the Kosslyn *et al.* (1998) study of mental rotation, summarized above. In that paper, we collapsed over counterbalancing orders for the two types of stimuli, Shepard & Metzler (1971) shapes and drawings of hands. We later realized that there were ample subjects in each group (those receiving hands stimuli first and those receiving Shepard & Metzler stimuli first) for an analysis of each group separately. The most interesting finding here concerned the areas activated while subjects performed the Shepard & Metzler task. When the Shepard & Metzler task was the first one that the subjects

received, we found that the inferior parietal and dorso-lateral prefrontal cortex were activated, as well as the right angular gyrus. However, when the subjects received this same task just after having performed the hands task, we found strikingly different results. Not only were over ten areas activated, but many of these areas had previously been found to be involved in motor control. Perhaps most strikingly, we found activation in the primary motor cortex itself, area M1. When we looked only at the data for the Shepard & Metzler task performed after the hands task, the data were in many ways similar to those for the hands rotation task itself.

Our interpretation of these, admittedly preliminary, results is that subjects could mentally rotate the Shepard & Metzler figures in two ways: they could imagine holding them and visualizing what would be seen if they twisted the forms around (i.e. a motor strategy, as has previously been documented (see, for example, Parsons 1987*a,b*; Wexler *et al.* 1998)), or they could imagine that an external force caused the rotation. Apparently, depending on what the subjects had just done, they were biased to adopt one or the other strategy. Note that this interpretation in turn leads to additional studies, including purely behavioural ones. For example, if subjects did in fact use different strategies, different patterns of response times and errors should arise (see Parsons 1987*a,b*); in addition, we would expect interference if subjects are asked to rotate a knob the opposite way to which they are rotating their images (see Wexler *et al.* 1998). Thus, the neuroimaging results serve to generate testable hypotheses, which in turn will lead to additional findings that will influence our interpretation of those results.

If our present interpretation of these neuroimaging findings is correct, these results are sobering: subjects have always been doing something before coming in the laboratory. Moreover, slight differences in the way in which a task is described or the instructions are couched could have large effects on how the subjects approach a task. If so, then we would expect to see many apparently contradictory results in the literature, where different findings emerge from the 'same' task. A walk through a poster session at a neuroimaging meeting will probably suggest that this fear may be well grounded. To address this issue, we must first discover whether the central assumption is correct: differences in context (including instructions) can lead to large differences in performance. If so, we must next take care to replicate precisely other laboratories' experiments before varying the design in any way; otherwise, we shall never know exactly which features of the results reflect which differences in the design (including context, in the broadest sense).

heading
(iii) *How does processing depend on the goal and the prior sequence of processing?*

Whether a particular process or structure will be invoked depends on the goal of processing and the specific processing that has already taken place. Thus, the last set of questions of this type addresses the principles by which specific sets of processes and structures come to work together. A crucial aspect of these questions focuses on temporal sequencing: When are processes evoked relative to each other? This type of question cannot be answered

with PET, and at present probably cannot be answered easily with fMRI. However, if data from individual trials are time-locked with the onset of stimulus, and enough trials are collected, it is possible that slight differences in the rise time of the haemodynamic response can be captured by fMRI. At present, the best way to collect data about the temporal course of large-scale neural processing (i.e. the activation of processes and structures of the sort discussed here) involves combining PET or fMRI with electrical or magnetic measures, such as MEG. However, the linkage between the different types of data is by no means straightforward. Electrical measures are notoriously poor in their spatial resolution, and thus it is difficult to bring them into registration with data from PET and fMRI. And although MEG has high spatial resolution and high temporal resolution, it is limited to detecting signals from the sulci (because of the local geometry of the neural architecture that gives rise to magnetic signals). Another promising technique is to use TMS to disrupt processing in specific areas (first identified via fMRI or PET) at specific intervals after a stimulus has been delivered (see, for example, Ganis *et al.* 1999).

This sort of research depends strongly on prior research of the sort outlined in the first section of this paper; without a road map of areas that underlie performance of a class of tasks, one cannot know where to look for temporal signals or where to disrupt processing. At the time of this writing, this approach is in its infancy.

4. CONCLUSIONS

I have tried to convey two overarching messages. First, the demonstration that a particular pattern of brain activity accompanies the performance of particular kinds of tasks is not, in and of itself, of great interest. Such data are only interpretable in the context of theories, which typically lead to specific hypotheses. Second, such theories can best be refined via a continual interplay between behavioural and neuroimaging studies, with results in each domain informing theory and further investigation in the other.

I have suggested that there are two general classes of questions that one can currently use neuroimaging to address: (i) how information processing gives rise to a particular ability, and (ii) when such processing is evoked. In addition, there are typically multiple sub-questions that one can ask within each general class. However, in every case there are crucial caveats. The interpretation of neuroimaging data is arguably more complex than the interpretation of behavioural data. One needs to be concerned not only about the nature of the task itself but also about the relation between performance and the underlying physiology. This job is all the more complex because we do not know whether increased rCBF (in PET) or signal strength (in fMRI) indicates excitation or inhibition, and we do not know the precise relation between these measures and cognitive work; moreover, this relation itself might be different for processes that have been used more or less frequently. Furthermore, the current resolution limitations prevent us from knowing how a given process actually works; we are stuck at the level of 'black boxes'.

Nevertheless, neuroimaging is a valuable tool for those interested in the nature of mental processes. It can provide converging evidence that has several strengths. First, it provides additional information about the underlying mechanisms themselves, about how different processes and structures work together. Second, such data are not easily explained as an artefact of experimental demand characteristics or the like; people rarely know how to manipulate voluntarily the level of rCBF in specific ROIs. Third, such data hold promise of providing a 'common language' for much of psychology. To the extent that we identify processing with brain regions, we then begin to see how a given process participates in conferring different abilities; this will provide crucial insights into what they have in common. Last, neuroimaging offers a potential bridge between psychology and biology, and then in turn to biophysics and genetics. This is no small feat; if we can find a way in which to benefit by even a small amount from the bounty in those fields, we shall benefit greatly.

To ask questions well, we must rely on theories. One source of inspiration about how information processing is implemented in the brain comes from animal models (see, for example, Lueck *et al.* 1989). However, such models are inherently limited; non-human animals lack the rich conceptual and linguistic processing of humans (and probably even have restricted emotional reactions compared with us). An additional source of inspiration are the effects of brain damage on human behaviour and cognition. In both cases, facts about neuroanatomy, particularly connectivity between areas, should have a crucial role in theory development. However, in the end we must pull ourselves up by our bootlaces. We must devise theories, test them, revise, and test again, keeping our eyes open for the unexpected all the while, just as in any other branch of science.

Neuroimaging is an enormously exciting field because of the possibility of new discoveries about fundamental relations between mind and brain, and the chance to make concrete what have hitherto been very abstract ideas. However, if we are to do this correctly, we must know what questions we are asking and just how far we can go in answering them with these techniques. Simply finding that certain areas of the brain are active when someone performs a task is not enough.

acknowledgements

I thank Giorgio Ganis, William Thompson and M. J. Wraga for helpful comments and suggestions on an earlier draft of this article. This research was supported by a grants NIA 5R01-AG12675-03, AFOSR F-49620-98-1-0334, and a grant from the John D. and Catherine T. MacArthur Foundation.

References

- REFERENCES**
- Andersen, R. A. 1987 Inferior parietal lobule function in spatial perception and visuomotor integration. In *Handbook of physiology*. 1. *The nervous system* (ed. V. B. Mountcastle), vol. 5, (*Higher functions of the brain*) (ed. F. Plum), pp. 483–518. Bethesda, MD: American Physiological Society.
- Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., Wu, J. & McGaugh, J. L. 1996 Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proc. Natl Acad. Sci. USA* **93**, 8016–8021.

- references
Cave, K. R. & Kosslyn, S. M. 1989 Varieties of size-specific visual selection. *J. Exp. Psychol. Gen.* **118**, 148–164.
- Checetto, D. F. & Saper, C. B. 1990 Role of the cerebral cortex in autonomic function. In *Central regulation of autonomic function* (ed. A. D. Loewy & K. M. Speyer), pp. 208–223. New York: Oxford University Press.
- Churchland, P. S. & Sejnowski, T. J. 1992 *The computational brain*. Cambridge, MA: The MIT Press.
- Damasio, A. R. 1985 The frontal lobes. In *Clinical neuropsychology* (ed. K. M. Heilman & E. Valenstein), pp. 339–402. New York: Oxford University Press.
- Desimone, R. & Ungerleider, L. G. 1989 Neural mechanisms of visual processing in monkeys. In *Handbook of neuropsychology* (ed. F. Boller & J. Grafman), pp. 267–299 Amsterdam: Elsevier.
- Felleman, D. J. & Van Essen, D. C. 1991 Distributed hierarchical processing in primate cerebral cortex. *Cerebr. Cortex* **1**, 1–47.
- Fox, P. T., Mintun, M. A., Raichle, M. E., Miezen, F. M., Allman, J. M. & Van Essen, D. C. 1986 Mapping human visual cortex with positron emission tomography. *Nature* **323**, 806–809.
- Ganis, G., Keenan, J., Pascual-Leone, A. & Kosslyn, S. M. 1999 Transcranial magnetic stimulation of primary motor cortex affects mental rotation. *J. Cogn. Neurosci.* **11** (Suppl., Abstract No. 70A), 86.
- Goldman-Rakic, P. S. 1987 Circuitry of primate prefrontal cortex and regulation of behavior by representational knowledge. In *Handbook of physiology*. 1. *The nervous system* (ed. V. B. Mountcastle), vol. 5 (*Higher functions of the brain*) (ed. F. Plum), pp. 373–417. Bethesda, MD: American Physiological Society.
- Goldman-Rakic, P. S. 1988 Topography of cognition: parallel distributed networks in primate association cortex. *A. Rev. Neurosci.* **11**, 137–156.
- Goodale, M. A. & Milner, A. D. 1992 Separate visual pathways for perception and action. *Trends Neurosci.* **15**, 20–25.
- Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch, P., Schapiro, M. B. & Rapoport, S. I. 1991 Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc. Natl Acad. Sci. USA* **88**, 1621–1625.
- Jolicoeur, P., Gluck, M. A. & Kosslyn, S. M. 1984 Pictures and language making the connection. *Cogn. Psychol.* **16**, 243–275.
- Kosslyn, S. M. 1980 *Image and mind*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. M. 1988 Aspects of a cognitive neuroscience of mental imagery. *Science* **240**, 1621–1626.
- Kosslyn, S. M. 1994 *Image and brain: the resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Kosslyn, S. M. & Koenig, O. 1995 *Wet mind: the new cognitive neuroscience*. New York: Free Press.
- Kosslyn, S. M., Cave, C. B., Provost, D. & Von Gierke, S. 1988 Sequential processes in image generation. *Cogn. Psychol.* **20**, 219–243.
- Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Chabris, C. F., Rauch, S. L. & Anderson, A. K. 1994 Identifying objects seen from different viewpoints: a PET investigation. *Brain* **117**, 1055–1071.
- Kosslyn, S. M., Thompson, W. L., Kim, I. J. & Alpert, N. M. 1995a Topographical representations of mental images in visual cortex. *Nature* **378**, 496–498.
- Kosslyn, S. M., Thompson, W. T. & Alpert, N. M. 1995b Identifying objects at different levels of hierarchy: a positron emission tomography study. *Hum. Brain Mapp.* **3**, 107–132.
- Kosslyn, S. M., Shin, L. M., Thompson, W. L., McNally, R. J., Rauch, S. L., Pitman, R. K. & Alpert, N. M. 1996a Neural effects of visualizing and perceiving aversive stimuli: a PET investigation. *NeuroReport* **7**, 1569–1576.

references

- Kosslyn, S. M., Thompson, W. L., Kim, I. J., Rauch, S. L. & Alpert, N. M. 1996a Individual differences in cerebral blood flow in area 17 predict the time to evaluate visualized letters. *Cogn. Neurosci.* **8**, 78–82.
- Kosslyn, S. M., Thompson, W. L. & Alpert, N. M. 1997 Neural systems shared by visual imagery and visual perception: a positron emission tomography study. *NeuroImage* **6**, 320–334.
- Kosslyn, S. M., DiGirolamo, G., Thompson, W. L. & Alpert, N. M. 1998 Mental rotation of objects versus hands: neural mechanisms revealed by positron emission tomography. *Psychophysiology* **35**, 151–161.
- Kosslyn, S. M., Pascual-Leone, A., Felician, 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 TMS. *Science* **284**, 167–170.
- Lang, P., Kozak, M., Miller, G., Levin, D. & McLean, A. 1980 Emotional imagery: conceptual structure and pattern of somato-visceral response. *Psychophysiology* **17**, 179–192.
- Lang, P. J., Ohman, A. & Vaitl, D. 1988 *The International Affective Picture System* (photographic slides). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.
- Levine, D. N. 1982 Visual agnosia in monkey and man. In *Analysis of visual behavior* (ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield), pp. 629–670. Cambridge: MIT Press.
- Lueck, C. J., Zeki, S., Friston, K. J., Deiber, M. P., Cope, P., Cunningham, V. J., Lammertsma, A. A., Kennard, C. & Frackowiak, R. S. 1989 The colour centre of the cerebral cortex in man. *Nature* **340**, 386–389.
- Luria, A. R. 1980 *Higher cortical functions in man*. New York: Basic Books.
- McClelland, J. L. & Rumelhart, D. E. (eds) 1986 *Parallel distributed processing: explorations in the microstructure of cognition. 2. Psychological and biological models*. Cambridge, MA: MIT Press.
- Mellet, E., Petit, L., Mazoyer, B., Denis, M. & Tzourio, N. 1998 Reopening the mental imagery debate: lessons from functional neuroanatomy. *NeuroImage* **8**, 129–139.
- Mishkin, M., Ungerleider, L. G. & Macko, K. A. 1983 Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* **6**, 414–417.
- Nyberg, L., McIntosh, A. R., Houle, S., Nilsson, L.-G. & Tulving, E. 1996 Activation of medial temporal structures during episodic memory retrieval. *Nature* **380**, 715–717.
- Parsons, L. M. 1987a Imagined spatial transformations of one's body. *J. Exp. Psychol. Gen.* **116**, 172–191.
- Parsons, L. M. 1987b Imagined spatial transformations of one's hands and feet. *Cogn. Psychol.* **19**, 178–241.
- Pascual-Leone, A., Grafman, J., Cohen, L. G., Roth, B. J. & Hallett, M. 1997 Transcranial magnetic stimulation. A new tool for the study of higher cognitive functions in humans. In *Handbook of neuropsychology*, vol. 11 (ed. J. Grafman & F. Boller), pp. 267–290. Amsterdam: Elsevier BV.

references

- Paus, T. 1996 Location and function of the human frontal eye field: a selective review. *Neuropsychologia* **34**, 475–483.
- Petersen, S. E., Van Mier, H., Fiez, J. A. & Raichle, M. E. 1998 The effects of practice on the functional anatomy of task performance. *Proc. Natl Acad. Sci. USA* **95**, 853–860.
- Petit, L., Tzourio, N., Orssaud, C., Pietrzyk, U., Berthoz, A. & Mazoyer, B. 1995 Functional neuroanatomy of the human visual fixation system. *Eur. J. Neurosci.* **7**, 169–174.
- Podgorny, P. & Shepard, R. N. 1978 Functional representations common to visual perception and imagination. *J. Exp. Psychol. Hum. Percept. Perform.* **4**, 21–35.
- Posner, M. I. & Petersen, S. E. 1990 The attention system of the human brain. *Annu. Rev. Neurosci.* **13**, 25–42.
- Price, C. J. & Friston, K. J. 1997 Cognitive conjunctions: a new approach to brain activation experiments. *NeuroImage* **5**, 261–270.
- Pylyshyn, Z. W. 1981 The imagery debate: analogue media versus tacit knowledge. *Psychol. Rev.* **87**, 16–45.
- Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M., Pardo, J. V., Fox, P. T. & Petersen, S. E. 1994 Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebr. Cortex* **4**, 8–26.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. 1976 Basic objects in natural categories. *Cogn. Psychol.* **8**, 382–439.
- Sergent, J., Ohta, S. & MacDonald, B. 1992 Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain* **115**, 15–36.
- Shepard, R. N. & Metzler, J. 1971 Mental rotation of three-dimensional objects. *Science* **171**, 701–703.
- Smith, E. E. & Medin, D. L. 1981 *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Thompson, W. L. & Kosslyn, S. M. 1999 Neural systems activated during visual mental imagery: a review and meta-analyses. In *Brain mapping. II. The applications* (ed. J. Mazziotta & A. Toga). New York: Academic Press. (In the press.)
- Treisman, A. M. & Gelade, G. 1980 A feature integration theory of attention. *Cogn. Psychol.* **12**, 97–136.
- Tye, M. 1991 *The imagery debate*. Cambridge: The MIT Press.
- Ungerleider, L. G. & Mishkin, M. 1982 Two cortical visual systems. In *Analysis of visual behavior* (ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield), pp. 549–586. Cambridge: MIT Press.
- Van Essen, D. C. 1985 Functional organization of primate visual cortex. In *Cerebral cortex* (ed. A. Peters & E. G. Jones), pp. 259–329. New York: Plenum Press.
- Wexler, M., Kosslyn, S. M. & Berthoz, A. 1998 Motor processes in mental rotation. *Cognition* **68**, 77–94.
- Wilson, F. A. W., O'Scalaidhe, S. P. & Goldman-Rakic, P. S. 1993 Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* **260**, 1955–1958.