On Cognitive Neuroscience

Citation	Kosslyn, Stephen Michael. 1994. On cognitive neuroscience. Journal of Cognitive Neuroscience 6(3): 297-303.
Published Version	doi:10.1162/jocn.1994.6.3.297
Accessed	February 18, 2015 7:16:24 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:3579184
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

On Cognitive Neuroscience

Stephen M. Kosslyn

■ Stephen M. Kosslyn is Professor of Psychology at Harvard University and an Associate Psychologist in the Department of Neurology at the Massachusetts General Hospital. He received his B.A. in 1970 from UCLA and his Ph.D. from Stanford University in 1974, both in psychology, and taught at Johns Hopkins, Harvard, and Brandeis Universities before joining the Harvard Faculty as Professor of Psychology in 1983. His work focuses on the nature of visual mental imagery and high-level vision, as well as applications of psychological principles to visual display design. He has published over 125 papers on these topics, co-edited five books, and authored or co-authored five books. His books include *Image and Mind* (1980), *Ghosts in the Mind's Machine* (1983), *Wet Mind: The New Cognitive Neuroscience* (with O. Koenig, 1992), *Elements of Graph Design* (1994), and *Image and Brain: The Resolution of the Imagery Debate* (1994). Dr. Kosslyn has received numerous honors, including the National Academy of Sciences Initiatives in Research Award, is currently on the editorial boards of many professional journals, and has served on several National Research Council committees to advise the government on new technologies. ■

JOCN: You played a major role in establishing the phenomenon of mental imagery as a tractable scientific problem. You started your work in the area of cognitive psychology but now have moved squarely into cognitive neuroscience. Why?

SK: The short answer is that facts about the brain allowed me to answer questions that seemed unanswerable using purely behavioral measures.

JOCN: And the long answer?

SK: My predecessors developed methods to study the *functions* of imagery, such as its role in memory and reasoning. I was interested in a different set of questions, concerned with the *structure* of the representations that underlie the experience of visual mental imagery. I consider these representations as types of data structures in an information processing system. In my original experiments, starting with one on image scanning in 1973, I used response time to try to infer properties of such representations. For example, I used response time as a kind of "mental tape measure" in the scanning experiments, with the goal of showing that mental image representations embody spatial extent. Introspectively, images seem to have pictorial properties, which seemed to make sense if the representation itself is a kind of spatial pattern. This type of representation would depict, rather than describe, the visual properties of an object or scene. If so, I reasoned, then people should require more time to shift attention farther distances across objects in their mental images (even when their eyes were closed). And this is just what happened: The farther people had to scan across an object to locate a named property, the longer it took.

The same year that the original scanning paper appeared, Pylyshyn published his critique of mental imagery. He argued that mental images are stored as "propositional" representations, no different in kind from the representations that underlie language. In his view, the pictorial properties of imagery that are evident to introspection are entirely epiphenomenal; they play no part in information processing. These properties are like the heat from a light bulb when one reads, which plays no role in the reading process. Thus began the so-called "imagery debate," which has kept me focused on imagery all these years.

The imagery debate was *not* about whether people experience mental images; all parties agreed that they do. It was about the nature of the underlying internal representations. Do the same types of representations underlie the experience of visual mental images and language, or is there something special about at least some of the representations used in imagery? I naively thought that the results from my scanning experiments spoke to this issue; they did, after all, show that a property of imagery that is evident to introspection—spatial extent—affects information processing. But this finding was easily explained in other ways. Some researchers argued that the visual properties of objects are represented as lists, and more time was required to iterate further down these lists; such lists preserve ordinal spatial relations, but do not depict information—they are not images. Some others argued that the instructions for the task led the subjects to use these types of representations (unconsciously) to mimic what they would do in the corresponding perceptual situation. At its heart, the problem was that the theories were too underconstrained. When faced with additional data, people could alter their notions about the properties of processes in order to preserve properties of their favorite representation. I found this state of affairs very frustrating. Presumably, there is a fact to the matter: When one has the experience of imagery, at least one of the underlying representations either has or does not have depictive properties.

So, why cognitive neuroscience? Neuroscientific information provided a way to ground this research, to remove some of the degrees of freedom that made it so easy to explain the behavioral results. When I did "dry mind" research, ignoring the brain, I argued that an image representation is like a pattern of points in an array in a computer. When I learned that multiple topographically mapped areas in the macaque cortex are used in visual perception, this made theorizing much more concrete and direct, and also provided grounds for making strong predictions: If one could show that at least some of these topographically mapped areas are active when one closes one's eyes and forms visual mental images, this would go a large part of the way toward demonstrating that image representations are depictive. And if imagery were disrupted when these areas are damaged, one could not argue that the representations they support are purely epiphenomenal. Moreover, Pylyshyn had raised a number of potential paradoxes; for example, does the "mind's eye" need a "mind's eye's brain"? And does the mind's eye's brain require its own mind's eye to "see" the images? Considering the roles of other areas that are connected to these topographically organized areas provided a handle on these issues. Turning to the brain not only helped me to characterize the questions, but invited additional approaches toward answering them—and these methods produced data that were more difficult to explain in other ways.

JOCN: Well, before we go into what your journey into brain science has taught you, would you care to define what you think the goals of cognitive neuroscience ought to be or might be? Your first thoughts simply reference some well known traditional neuroscience work. Is cognitive neuroscience a new intellectual discipline or simply traditional neuropsychology dressed up in a new phrase?

SK: Cognitive neuroscience is a good illustration of how the whole can be more than the sum of its parts. In my view, cognitive neuroscience is an interdisciplinary melding of studies of the brain, of behavior and cognition, and of computational systems that have properties of the brain and that can produce behavior and cognition. I don't think of cognitive neuroscience as the intersection of these areas, as the points of overlap, but rather as their union: It is not just that each approach constrains the others, but rather that each approach provides insights into different aspects of the same phenomena.

When you ask about "traditional neuropsychology," I assume that you don't mean the early work by clinicians that was designed to detect brain injury; this work was extremely empirical, and not aimed at understanding the underlying mechanisms. The more interesting comparison, I think, is to "cognitive neuropsychology." Both cognitive neuropsychology and cognitive neuroscience make use of theory developed in cognitive psychology and cognitive science to characterize the nature of the behavior or cognitive process to be studied. And both enterprises want to specify how information processing occurs; indeed, both sets of researchers often rely on computational models. Moreover, both enterprises exploit tasks and methodologies that have been developed in cognitive psychology and cognitive science (and these tasks and methods are often more sophisticated than those used in traditional neuropsychology). The major contrast between cognitive neuroscience and cognitive neuropsychology is revealed by the different nouns in their names. Cognitive neuroscience is an attempt to understand how cognition arises from brain processes; the focus is on the brain, as the term "neuroscience" implies. We don't want to separate the theory of information processing from the theory of the brain as a physical mechanism. Cognitive neuropsychology, at least as characterized by Caramazza, Shallice, and others, focuses on the functional level per se. They want to understand information processing independently of properties of the wetware itself.

A complete cognitive neuroscience theory would specify more than just the component processes and principles of their interaction. In addition, it would specify how each process is instantiated in the brain, and how brain circuits produce the input/output mappings accomplished by each process. This understanding would extend down to individual types of receptors, channels, and ultimately to the genes. Given these goals, it seems clear that cognitive neuroscience must move closer to neurobiology. But it will not simply become neurobiology: Cognitive neuroscience adds methods and techniques to study, and conceptualize, how the brain gives rise to cognition and behavior.

JOCN: For some, the componential nature of the new cognitive psychology translated nicely into the neurologic clinic where bizarre dissociations are the rule. Perhaps, it was felt, processing modules could be selectively hit with brain lesions and therein provide support for a cognitive formulation. Yet would you not agree the *objective* of a mature cognitive neuroscience would be to ascertain the algorithms active in translating structural/physiological data into psychological function.

SK: Yes, that's part of the objective. We want to understand not just the component processes, but also the details of how neurons actually compute these functions—the algorithms, if you will.

JOCN: Isn't this what David Marr had in mind? What would you say is Marr's greatest contribution, looked at with the cold lenses of hindsight some dozen years after his death?

SK: I think cognitive neuroscience owes an enormous amount to David Marr. Marr provided the first concrete, well-worked-out example of how one could rigorously combine neuroscientific, computational, and psychophysical findings and concepts. He provided an illustration of how the different sorts of information could illuminate a single problem, providing insights into different aspects of it. Even though the details of his particular theory of visual processing may not stand the test of time, his style of thinking and approach are nothing short of brilliant. In my view, one of Marr's best ideas is his conception of a "theory of the computation," which has received surprisingly little attention. He was unhappy with the tendency in Artificial Intelligence research to make up theories purely on the basis of intuition, and wanted theories to be rooted in careful logical analyses and empirical investigation. Marr argued that one should develop a theory of the computation whenever one proposes a particular decomposition into processing components. Such a theory rests on a detailed analysis of what information processing problems must be solved in order for a system to be capable of having certain abilities; the abilities are determined empirically, from studies of normal cognition and behavior and studies of cognition and behavior following brain damage. Once one has a theory of the goal of a processing component or set of components, what they're for, one then is in a position to theorize about the specific representations and algorithms that are used. I can't possibly do justice to these ideas here, so let me simply recommend strongly that people go back and read Marr's book, if only to understand his style of thinking. Too much of his good advice has been neglected by contemporary "connectionist" modelers, who often seem to make up theories at the level of the algorithm as they go along.

JOCN: Does Marr's approach guide you?

SK: I try to develop "poor man's versions" of theories of what is computed. I simply don't have Marr's gift for seeing how to formalize vaguely specified problems. In my new book, *Image and Brain,* I've tried to use Marr's approach in a qualitative way, and even this seems preferable to relying solely on intuition and attempts to explain empirical results.

JOCN: OK, let's take the problem of mental imagery and go through how the brain side of the story has evolved over the last 10 years. First, what has the lesion work instructed us about imaginal processes?

SK: Two main messages emerge from the lesion work: First, imagery and like-modality perception share many common mechanisms, even though they do not rely on identical mechanisms. One often sees corresponding deficits in imagery and perception (such as unilateral visual neglect, as documented by Bisiach and his colleagues), but also can find patients who have intact imagery and deficient perception (e.g., as documented by Behrmann and her colleagues) and vice versa (e.g., as demonstrated by Charcot, Brain, and others many years ago). The finding that imagery shares many mechanisms with perception is very important because it is much easier to understand perception than to understand imagery: Not only is perception rooted in observable stimulus events (which can be experimentally manipulated and correlated with psychological events), but also we have very good animal models of our perceptual systems and hence have come to understand the underlying neural systems in considerable detail. We can "piggyback" on this understanding when developing theories of imagery.

Second, results from lesion studies have shown that imagery is not a single process. For example, the "what vs. where" distinction that Ungerleider and Mishkin introduced in visual perception also extends to imagery (e.g., as demonstrated by Levine, Calvanio, and ·Farah). Moreover, different imagery abilities can be selectively disrupted by brain damage. For example, my colleagues and I have described patients who can generate and maintain images (at least of the types we tested) but have difficulty rotating objects in images.

JOCN: Are you really comfortable with these conclusions? Isn't the lesion method full of difficulties? For example, couldn't the dissociation you mention simply be reflecting task difficulty? Surely it is more difficult to rotate an image and given that, the brain damage itself rears its ugly head?

SK: The lesion method, like all others, has potential problems and has to be used with care. For example, as you note, more difficult tasks will not be performed as well by patients with brain damage—and so a dissociation may say nothing about the existence of distinct processing components. But this is not an insurmountable problem. One way to deal with it is to design tasks that are equated for difficulty. The most straightforward way to do. this is to pretest age- and education-matched control subjects, and adjust the materials until these subjects require the same mean time and have the same mean error rates in the tasks. Another response to this sort of possible problem is to obtain a "double dissociation," to find two patients with the opposite pattern of deficits. However, even when one does find a clean double dissociation, such a result is not airtight evidence for the existence of separate processing components; Shallice's book has a nice discussion of the applicability of the logic of double dissociation, and in 1992 Intriligator and I published a paper (in this journal) that touched on this topic.

JOCN: Aren't there better approaches? For example, can a cognitively impoverished disconnected right hemisphere carry out mental rotation tasks or can only the cognitively superior left hemisphere?

SK: As far as I can tell, all methods have their drawbacks, and their strengths. Hemispheric dissociations are a valuable source of converging data, but they too are sometimes ambiguous. For example, there are reports that patients with unilateral left- and unilateral right-hemisphere lesions have deficits in mental rotation, and there are divided-visual-field studies of normal subjects that report that the right hemisphere is better than the left hemisphere, that the left hemisphere is superior, or that both hemispheres are equally effective. I don't find this surprising: Any complex task is likely to be performed by many component processes working in concert, some of which may be more effective in one hemisphere and some of which may be more effective in the other hemisphere—and depending on the precise nature of the task, different components may contribute more or less to the overall processing, leading to hemispheric differences in performance.

JOCN: So, what are the strengths of the lesion approach?

SK: I find lesion data particularly useful for testing predictions: If one posits that the posterior parietal lobe does function X, patients with damage to this area had better show deficits in tasks that rely on this function. Similarly, if one claims that a particular region is the seat of a specific processing component, then damage to other regions should not affect that function (when factors such as overall activation level and diaschesis are controlled). Lesion work can play a critical role in telling one whether a specific area is necessary and/or sufficient for a specific type of processing.

In addition, selective deficits following brain damage have enormous heuristic value; even though these deficits do not always reflect the loss of individual processing components in isolation, they sometimes may. And thus they can serve as a useful source of hypotheses, they can inform one's theory of what is computed. Harking back to Marr, his theory of how shapes are stored in long-term memory was influenced by Warrington's finding that some types of brain-damaged patients could not identify objects seen from unusual points of view. There is no guarantee that the hypothesis is correct, of course (and I think Marr was off the mark in this case), but that's not the point: The dissociations following brain damage can lead one to formulate interesting hypotheses, which in turn lead to further empirical investigation.

In general, lesion data, like all other sorts, are best used as one source of converging evidence. There are lots of *potential* problems with any method. This doesn't mean that these potential problems are actual problems in any specific case.

JOCN: Converging evidence is important, to be sure. However, neuroscience often seems to depend on double dissociation as the solid test of theories about cognitive function. Perhaps convergent evidence is called for only when a double dissociation cannot be found. For example, it would be hard to find a patient that can rotate an image they can't generate!

SK: Actually, I would predict that there should be patients who can rotate objects in images but have difficulty generating them; in the typical rotation task, the stimulus remains in view, and the task does not require activating information stored in long-term memory (in Cooper and Shepard's famous tasks, the subjects usually are simply asked whether a visible figure is facing normally or is mirror reversed, regardless of its angular orientation). In any case, double dissociations are only one of a number of sources of evidence that can converge to support a particular theory of cognitive processing.

JOCN: Well, this notion of converging data is a version of the meta-analysis approach, is it not? Who was it that said if you see a pile of shit, you know there must be a pony in there somewhere? Psychological processes are full of probabilistic events that allow meta-analysis to work. When it comes to biological processes, however, the approach seems inappropriate. Either you know something or you don't. Either something is built in a certain way or it is not.

SK: I don't think of converging evidence as a version of the meta-analysis approach. That approach is most commonly used to find statistical significance over a set of individual studies, each of which may have reported nonsignificant findings. In the approach I advocate, each individual finding should be statistically significant. That's not at issue. What's at issue is how to ***interpret*** the individual results. And it is here that convergent evidence is so important, given that results from any given method usually are open to alternative interpretations. To say that "you know something or you don't" is correct if we define "something" as a specific result; one knows that a patient with lesion X, or with information sent only to hemisphere Y, does this-and-that well but not this-and-the-other-thing. The results themselves do not imply that one knows that a specific process exists or that a particular sequence of information processing takes place; it's up to the theorist to interpret the data.

In my view the convergent evidence approach does two things for you. First, it helps you figure out whether a given finding is due to some kind of artifact or methodological problem. There are plenty of such possible snags with *any* method, but these are *potential* problems and need not necessarily bedevil a given experiment. The best way to know whether to take a specific set of results seriously, in my view, is to see whether it lines up with results from other methods (which also have potential problems, but different ones).

Second, the converging evidence approach does more than simply validate the different methods. It fleshes out the nature of the phenomena to be studied, and provides insights into different aspects of the underlying mechanisms. A good convergent evidence approach uses the results from one type of study to guide other types of studies. For example, results from our recent fMRI studies suggest that only some subjects activate primary visual cortex during imagery, although the other subjects do activate other regions of the occipital lobe. This result leads to an hypothesis about processing: Do subjects who activate primary visual cortex have more vivid images than those who do not activate this region? If so, they should be able to answer questions that require high-resolution images better than subjects who don't activate primary visual cortex, but there should be no difference if low-resolution imagery is all that is necessary. Similarly, this line of thinking leads to asking whether primary visual cortex is necessary for high-resolution images, which could be tested by examining patients who have selective damage to primary visual cortex, but little or no damage to circumstriate regions; will they form fuzzier images than patients with equivalent damage to other parts of the brain? And so forth.

So, different methods are used to ask about different facets of a problem. I think it is important to keep in mind that depending on what particular question you ask, different things count as answers. And depending on the kind of answer you seek, different methods will be more or less appropriate. If you want to know whether a specific area is involved in processing, a brain-scanning study makes sense. If you want to know whether this area is necessary for such processing, a lesion study makes sense, and so forth.

JOCN: But the convergent evidence approach breaks down if the results don't line up, doesn't it? You mention brain-scanning results. What we have with PET results is, at this point, a set of findings. It now appears impossible to activate, using PET, the frontal eye fields. If PET is tracking activity, how could that be? It now appears the hippocampus is not activated during memory tasks. How could that be? A recent review of language studies finds each investigator has activated different cortical areas for phonological and semantic processing. How could that be?

SK: The convergent evidence approach doesn't imply that the results necessarily must converge . . . only that they will make sense if you do the right experiments and your theory guides you to look for the right characteristics of the data. If you don't design a task properly to engender a specific type of activation, you won't see the brain footprints of that type of processing. And these footprints need not be consistent activation of a single locus. They could be activation of a pattern of areas, of any *k* of *n* possible areas, and so forth; my own view is that individual areas can be characterized as having specific functions that will reliably be reflected by PET activation, but that's not the only possibility. With PET the situation is particularly tricky because most PET work involves subtracting blood flow in one condition from blood flow in another. Depending on the nature of the baseline task, different patterns of activation will be evident.

JOCN: If there are constraints in understanding lesion data and hemisphere data, I suppose there are constraints in interpreting PET data. You have recently jumped into the PET arena and have published a fascinating report that visual imagery involves primary visual cortex. Give a quick synopsis of that study and tell us what you think the data can mean.

SK: The PET research on imagery that we've conducted, in conjunction with Nat Alpert and the MGH group, has centered on resolving the "imagery debate." As I mentioned at the outset, this debate focused on the nature of the internal representations underlying the experience of imagery. Specifically, when one experiences a visual mental image, is a picture-like "depictive" representation being processed? My colleagues and I argued that—miraculously!—introspection can sometimes reveal properties of the functional representations. To investigate this issue, we showed that topographically mapped visual cortex is activated when one forms visual mental images, even if one's eyes are closed. In addition, we found that spatial properties of images systematically affect the activation in these areas: When subjects visualized letters so that they seemed to subtend large visual angles, the centroid of activation shifted toward the anterior portions of this topographically mapped area, relative to when the subjects visualized letters so that they seemed to subtend small visual angles. In fact, the coordinates of these centroids were reasonably close to where they should be, based on the estimated "size" of the imaged letters (using techniques I developed in the 1970s to estimate the "visual angle" subtended by imaged objects). I am now fairly confident that the activated area in medial occipital cortex probably is area 17, especially given results from later work with functional magnetic resonance imaging (fMRI, e.g., from Ogawa and Tank, from Le Bihan and Turner, and from a collaboration we have with Belliveau at the MGH); this technique allows more precise localization within a single individual.

What do such findings mean? It is well known that most areas in the visual system (of the macaque) that have afferent connections to other areas also receive efferent connections from them; the connections are reciprocal. I argue that visual information is stored in a type of compressed code, and that imagery occurs when visual memories cause activation to flow backward in the visual system, along the efferent connections, to reconstruct a pattern in topographically organized cortex. By so doing, the shape, color, and spatial properties of objects are made accessible for additional processing. For example, your visual memory of a German Shepherd dog is probably stored in inferior temporal cortex using some kind of population code, which specifies shape by a vector defined over a large set of neurons with complex response properties. If I ask you whether the dog's ears are pointed, or whether the ears sit on the top or sides of its head, or whether they protrude above the top of its head, you will probably generate a visual mental image to reconstruct the actual spatial layout. Once you've generated the image, you can "take a second look" and reinterpret information that was only implicit in your stored memories. As this view implies, we did in fact find activation in inferior temporal cortex and a variety of other areas that presumably are used in generating and interpreting visual images.

JOCN: So, are you bothered by the fact that there are findings suggesting mental imagery goes on in other cortical areas such as the frontal lobes?

SK: Not at all; this is exactly as we predicted. We have argued that imagery involves depictive representations, which occur in topographically mapped regions of the occipital lobe, but imagery also involves nondepictive long-term memory representations (which we think are stored in the inferior temporal lobe) and lots of processing (including in frontal areas) to generate and use images. As Marr argued, the brain apparently implements many, very specialized "computations," which may be carried out in different regions. Any complex activity, such as imagery, perception, or memory, is likely to be accomplished by a host of relatively simple computations that work in concert. Our PET results show that a *system* of areas is involved in carrying out imagery. So, for imagery, we find that areas in the frontal lobes that are used to direct attention to key aspects of visual stimuli are also activated when one generates images. We hypothesize that high-resolution images are generated by activating visual memories of individual parts or properties, and "placing" them in the appropriate relative locations. This process, we argue, relies on the same machinery used to shift attention to search for a distinctive part of an object during perception. Other imagery activities, such as mental rotation or image maintenance, would use other combinations of simple component processes. So, from this perspective we expect different tasks to result in different patterns of brain activation.

JOCN: Given the large number of assumptions, is this technology really useful for cognitive neuroscience?

SK: All methods rely on lots of assumptions, so that fact alone can't be a criticism of PET or fMRI per se. It's still too soon to know what the critical assumptions are for these techniques; I suspect that the best way to find out is to keep using the techniques and vary various parameters, discovering what is and is not important. Is the technology useful? Depending on the question one asks, different things count as answers. If one wants to know whether two tasks rely on the same processing, then showing that the same pattern of brain activation occurs during both can help one to answer that question. If one wants to know whether two processes are the same or different, then finding separate patterns of activation for them can answer the question. In my view, the new scanning technologies are likely to play an even greater role in cognitive neuroscience as we begin to characterize what distinct areas of the brain do (e.g., anterior cingulate cortex, area 46, etc.); once we have characterized what an area does, we then can start to draw inferences about how subjects perform a task that activates that area. Given that the area is activated, one has evidence that a specific process is used to perform the task. This sort of reasoning is not always going to be simple or straightforward, however, because a number of different processes may turn out to be supported by the same tissue, or the function of a given area may turn out to depend in part on what other areas are doing, but the more we understand about what an area of the brain does, the more we can learn about a task that activates that area. But again, let me stress, I think convergent evidence is the way to go; there is no Royal Road to understanding how the brain gives rise to the mind.

JOCN: PET, then, will play a greater role in sharpening ideas about cognitive models of imagery or memory or whatever. I guess you don't see it playing a role in actually instructing the neuroscience side of the equation, namely the physiological mechanism active in enabling a cognitive state. After all, when activation is detected, it is not at all clear whether it reflects inhibition or excitation at the synaptic level.

SK: I'm no expert on the physiology of PET, but I think it's too soon to foreclose any specific use of it. PET may well turn out to be a tool that can be used to determine whether a particular activation reflects net inhibition or excitation. I would argue that the problem of characterizing neural activity will be solved only by developing and testing specific theories. In cognitive neuroscience, we are trying to develop theories of what sets of neurons do. In my view, as we understand how specific components of the functional architecture are implemented in the brain, we will necessarily come to understand more about the neural substrate. Our ignorance of whether activation reflects net excitation versus inhibition pales besides our ignorance of what are the consequences of a local pattern of activation for processing in the system as a whole. The ambiguous nature of activation is actually much worse than you note: It's not simply that we don't know whether the activity reflects net inhibition or excitation at the synaptic level, it's that we don't know what the area is doing (is it activating stored information? releasing some other area from inhibition? selectively activating a process implemented elsewhere? transforming input into a different kind of output? etc.) and we don't know how the area is carrying out this computation. PET is one method that will help us to answer such questions, and in so doing we will understand more about the brain itself. PET will not be the only tool to advance our knowledge of neural activity, and probably couldn't do it alone, but it will be one source of convergent evidence.

JOCN: But for the cognitive modeler, the details of what is happening at a synaptic level are not important. Right?

SK: To the contrary, in my view the two enterprises—understanding cognitive processing and understanding the neural substrate—mutually inform each other; they are different facets of the same problem. It is clear that further insights about neurophysiology and neuroanatomy inform the cognitive end (for example, Rockland's recent findings of direct connections from area TE to area V1 have clear implications for theories of imagery), and as we come to understand the nature of cognitive processing, that should inform theories of the neural substrate (e.g., my view of what V1 does has been changed by our PET results). A more detailed understanding of the neural activity that underlies a specific pattern of activation will aid cognitive modeling. Indeed, someone wanting to build a "realistic" neural network model will very much want to know patterns of excitatory and inhibitory interactions at the synaptic level; and even questions about the nature and organization of processing subsystems will be easier to answer as we know more details about the neural events that produce specific activation and the specific anatomic connections among local portions of the brain.

JOCN: So, what's the next step? You've been working on visual mental imagery for over 20 years now; what do you see for the next 20 years?

SK: We've begun to make a dent in understanding the mechanisms that allow us to produce and use mental images, but it is clear that this is only a dent. My work has become increasingly focused on understanding the role of specific content in directing and modulating processing. For example, a major question that has received too little attention concerns the role of imagery in emotion. Why do vivid images often accompany highly emotional memories? What roles do these images play? How does the imagery system that we've begun to characterize interact with the neural systems that underlie emotion? My current goal is to use PET and fMRI to try to understand (at a relatively coarse level) the circuitry that causes one's palms to sweat when one visualizes a threatening scene (e.g., teetering on a narrow trail etched into the side of a very steep mountain). As part of this effort, I would like to understand the role of imagery in classical conditioning. Thirty years ago this would have seemed a very odd juxtaposition indeed, but it is now possible to study such questions—and perhaps even to begin to answer them!

JOCN: Thank you.

Reprint requests should be sent to Stephen M. Kosslyn, Department of Psychology, 832 William James Hall, Harvard University, Cambridge, MA 02138.