Two types of image generation: Evidence from PET

STEPHEN M. KOSSLYN

*Harvard University, Cambridge, Massachusetts
and Massachusetts General Hospital, Boston, Massachusetts*

WILLIAM L. THOMPSON and KATHERINE E. SUKEL
Harvard University, Cambridge, Massachusetts

and

NATHANIEL M. ALPERT
Massachusetts General Hospital, Boston, Massachusetts

Is there more than one method whereby people can generate visual mental images? Participants generated images after learning patterns in two ways. In one condition, they memorized descriptions of how segments are arranged to form patterns; in another, they memorized segments and mentally amalgamated them into patterns. In both conditions, identical stimuli cued them to form images while brain activation was monitored using PET. Comparison of the two imagery conditions revealed different activation between hemispheres when images were formed after patterns were learned by mentally combining segments versus when images were formed from stored verbal descriptions. Thus, images can be generated in at least two ways. However, this laterality difference was subtle; the majority of brain areas were activated in common across conditions. This overall similarity is counter to what would be expected if image generation is simply perceptual exploration in the absence of appropriate stimuli, as is posited by perceptual activity theory.

One of the major advances of cognitive neuroscience is that mental functions that previously were regarded as unitary and undifferentiated have been shown to have a complex underlying structure. Visual mental imagery is one such function. Previous research has indicated that it is possible to decompose imagery into various processes, such as image generation (the process of creating an image on the basis of stored information), image inspection (interpreting patterns in images), and image transformation (changing the appearance of visualized patterns; see, e.g., Farah, 1984; Kosslyn, 1994; Kosslyn, Brunn, Cave, & Wallach, 1984; Kosslyn et al., 2004). In this article, we push this line of research one step further, and show that even a single process—image generation—can be further decomposed. We use positron emission tomography (PET) to provide evidence that visual mental images can be generated using two distinct methods.

The process of generating images is constrained in part by the nature of the information that is used to form the images. People typically look at several portions of

an object when examining it, which implies that separate representations of parts and characteristics of objects are encoded. If a visual mental image is later to be formed, the stored parts and characteristics must be properly amalgamated. Kosslyn, Maljkovic, Hamilton, Horwitz, and Thompson (1995) have argued that there are two ways in which parts can be arranged to form images. On the one hand, they can be arranged by using *categorical* spatial relations, which specify equivalence classes (e.g., *above*); on the other hand, they can be arranged by using metric information that specifies (with precise distances) exactly how parts are juxtaposed (see also Kosslyn, Thompson, Gitelman, & Alpert, 1998). Kosslyn et al. (1989; Kosslyn, Maljkovic, et al., 1995; Kosslyn et al., 1998) showed that categorical spatial relations are encoded more effectively in the left cerebral hemisphere, whereas *coordinate* spatial relations (which specify metric information relative to an origin) are encoded more effectively in the right cerebral hemisphere; this dissociation provides evidence that the two types of spatial relations are encoded by distinct mechanisms (see also Chabris & Kosslyn, 1998; Hellige & Michimata, 1989; Jager & Postma, 2003; Kosslyn, Chabris, Marsolek, & Koenig, 1992; Laeng, 1994; Laeng, Chabris, & Kosslyn, 2002; Slotnick, Moo, Tesoro, & Hart, 2001). If two types of spatial information are encoded and stored, in principle, each type could be used subsequently to reconstruct the spatial arrangement of the parts in a mental image;

This research was supported by ROLE Grants REC-0106760 and REC-0411725 from the National Science Foundation and by Grant R01 MH60734 from the National Institutes of Health. We thank Avis Loring and Steve Weise for technical assistance in conducting the study. Correspondence concerning this article should be sent to S. M. Kosslyn, Department of Psychology, Harvard University, 832 William James Hall, 33 Kirkland Street, Cambridge, MA 02138 (e-mail: smk@wjh.harvard.edu).

and if different processes are required to use the two types of spatial representations, we should find some brain areas more active during one type of processing and others more active during the other type of processing (see Kosslyn, 1987, 1994).

Kosslyn, Maljkovic, et al. (1995) used a divided visual field technique to test the hypothesis that visual mental images can be constructed in two ways, one of which employs categorical spatial relations to arrange parts and one of which uses coordinate spatial relations to arrange parts. To test this hypothesis, they adapted a task devised by Podgorny and Shepard (1978). In this task, participants are instructed to visualize a block letter within a grid or within a set of four corner brackets (which delineate a rectangular space). An X mark appears within the grid or the space defined by the brackets, and the participants decide whether the X mark would have fallen on the letter, had it been present. The results from Experiments 1 and 2 of Kosslyn, Maljkovic, et al. are of most relevance here. In Experiment 1, the participants were initially given descriptions of how segments are arranged to form a shape; they were asked to memorize the descriptions well enough to be able to visualize the corresponding pattern. The descriptions forced the participants to use categorical spatial relations when building up an image of the pattern. In Experiment 2, the participants viewed a sequence of segments, one at a time, on a screen, "mentally glued" them into a single pattern, and were asked to memorize this pattern. In order to construct an image of the composite pattern, the participants needed to remember exactly where each segment was, which required coordinate spatial relations. After memorizing the different types of stimuli, the participants were cued to visualize one of the patterns within a set of four corner brackets and to decide whether the pattern would cover an X mark that was presented within that region.

The results from these experiments were straightforward. If the participants had memorized a description, they could perform the task more accurately when the stimuli appeared in the right visual field and, hence, were encoded initially by the left hemisphere (LH). In contrast, if the participants had memorized a sequence of individual segments, they could perform this task more accurately when the stimuli appeared in the left visual field and, hence, were encoded initially by the right hemisphere (RH). The same stimuli—X marks within sets of four corner brackets—were used in both experiments, and hence, the results were produced by the way the stimuli were processed.

The claim that mental images can be generated in two ways, which draw preferentially on the LH or the RH, has proven controversial. The traditional view is that imagery is an RH function (see, e.g., Ley, 1979), but this view was challenged by Farah (1984) on the basis of a review of the published reports of case studies of brain-damaged patients. Moreover, Farah (1986) reported divided visual field studies that suggested that normal people form images more effectively when the LH is cued. However, Sergent (1990) took issue with Farah's

analyses of the case studies and argued that the data are in fact equivocal. In addition, Kosslyn, Holtzman, Farah, and Gazzaniga (1985) showed that each of the isolated hemispheres of split-brain patients could generate images but appeared to do so in different ways. Sergent (1989) then demonstrated that normal participants apparently generate images in some tasks using RH processes. Kosslyn (1987) summarized divided visual field studies that suggested that participants could adopt either of two strategies, which drew primarily on one or the other hemisphere. More recently, Kosslyn and Thompson (2003) reviewed 59 studies of visual mental imagery (all requiring the participants to generate images) and found no evidence that one hemisphere is used preferentially during imagery. In short, these waters are muddy (cf. Tippett, 1992).

In the present study, we essentially replicated Experiments 1 and 2 of Kosslyn, Maljkovic, et al. (1995), as summarized above, but now used PET to monitor regional cerebral blood flow as the participants performed the task. Before the *verbal condition*, they memorized verbal descriptions of where segments were located in the grid; before the *amalgamation condition*, they saw individual segments and mentally amalgamated them into a single shape. The participants later were cued to visualize the resulting image and to determine whether an X would have fallen on the pattern. If, in fact, people use categorical spatial relations to assemble parts in one case and coordinate spatial relations to assemble parts in the other, we would expect differences in the patterns of brain activation in the two situations. But more than that, we would expect to find different activation in the cerebral hemispheres when the two conditions are directly compared: Images formed on the basis of categorical spatial relations representations should selectively draw on processes in the LH, whereas images formed on the basis of coordinate spatial relations representations should selectively draw on processes in the RH. Such results would provide strong evidence that there are, in fact, two ways to generate images.

METHOD

Participants

Twenty-three right-handed male participants (mean age, 21 years, with a range of 18–31 years) responded to a newspaper advertisement and served as paid volunteers. All were university students or young professionals from the Boston area who reported normal or corrected-to-normal vision and general good health. One volunteer terminated his participation due to claustrophobia, and 6 participants performed the task at a chance level of performance, so that we could not infer that the appropriate processing had taken place; hence, the PET data from these participants were not analyzed. The remaining 16 participants were randomly divided into two groups (with mean ages of 22 years, range of 19–23 years, vs. 21 years, range of 18–27 years). None of the participants was aware of the hypotheses at the time of testing. All the participants provided informed consent and were tested in accordance with national and international norms governing the use of human research participants. The research was approved by the appropriate Harvard University and Massachusetts General Hospital Institutional Review Boards.

Materials

The materials were identical to those used in Experiments 1 and 2 of Kosslyn, Maljkovic, et al. (1995). As is illustrated in Figure 1, a *donut grid* was produced from an 8×9 grid by removing the lines from a 4×5 region at the center (thus producing a hole in the middle); this stimulus was used during the learning phase. The donut hole measured approximately 2.3 cm horizontally by 3.0 cm vertically, or $2.6^\circ \times 3.4^\circ$ of visual angle from the participants' vantage point of 46 cm. This hole defined the field in which the participants were asked to visualize each stimulus figure and was used in the study phase of both the verbal and the amalgamation conditions. The grid lines in the stimulus were provided in the verbal condition to serve as guidelines to help the participants mentally place all the segments of each image in the proper position.

In addition, we created 32 sets of test stimuli by producing 4×5 grids of the same size as the donut hole in the study stimulus and then removing all of the lines except for four brackets at the corners. A single X mark was placed within the space delineated by the brackets in each of these stimuli; if the grid lines had been present, the X would have precisely connected the corners of a single cell. The X mark appeared equally often in each quadrant of the space within the brackets. Moreover, each stimulus was paired with a figure, and the Xs were positioned so that half would have fallen on the figure had it been present and half would have fallen adjacent to it. Furthermore, half of the Xs that would have fallen on it were placed on the first or second segment (as the participants learned them), and half were placed on the penultimate or final segment.

In the verbal condition, we created four descriptions of patterns, which specified the relative locations of segments. The four figures had novel shapes, and thus the participants had to memorize each description verbatim. An example of one of these descriptions is as follows. "A vertical bar extends from the bottom left corner upward to just past (above) the bottom left corner. Attached to the right end of the first bar is a horizontal bar extending rightward to the end of the row. Attached to the top end of the second bar is a vertical bar extending upward to the top right corner."

In the amalgamation condition, we created four sets of individual segments for the same four patterns as those described in the verbal condition; each segment was shown on the screen and then removed before the next segment in the sequence was presented. Each segment was identical to one that was specified by a phrase in the verbal condition. For example, the first phrase in the description above ("a vertical bar extends from the bottom left corner upward to just past [above] the bottom left corner") corresponded to an amalgamation condition study stimulus where the donut grid featured a vertical segment filling only in the single grid cell in the bottom left corner. All the instructions and stimuli were presented on a Macintosh computer (Apple Computer, Cupertino, CA), with a 7-in. black-and-white screen display; this computer was mounted on a gantry, 46 cm from the participant's eyes, and presented stimuli using a modified version of the MacLab program (Costin, 1988).

Task Procedure

All the participants took part in three conditions: a baseline, a verbal image generation, and an amalgamation image generation

Figure 1. Illustration of the stimuli used during verbal learning (top) and amalgamation learning (bottom). The arrows separate information presented on separate sheets or computer screens. Two figures were used in each of the learning conditions, and the figure/condition pairings were counterbalanced across participants.

condition. We arbitrarily paired the figures into two sets and presented one set in each of the two experimental conditions; thus, a given participant memorized two stimuli using one method and two using the other method. The baseline was always presented first, but we counterbalanced the order of experimental conditions (verbal or amalgamation, after baseline), figure (which figure of the pair would be learned first), set (which pair of stimuli would be used in which condition), and the order of the segments in the figures (top to bottom or bottom to top). One participant was tested for each combination of these orderings, creating the complete design: 2 (condition order) $\times 2$ (figure order) $\times 2$ (set used per condition) $\times 2$ (segment order) = 16 participants.

Baseline condition. The general purpose and procedure were explained to each participant prior to testing. After signing the informed consent form, the participant entered the tomograph and was fitted with a thermoplastic custom-molded face mask (TruScan, Annapolis, MD) to keep the head as still as possible during scanning. Instructions for the task were presented on the computer's screen. The participants were asked to fixate on an asterisk in the center of the screen. After 1 sec, the asterisk disappeared and was replaced with either the number 1 or the number 2 for 500 msec. Immediately following this, the asterisk appeared for another 500 msec and was then replaced by a set of brackets-with-X for 150 msec. The asterisk then reappeared and remained in view until a response was made. The participants were asked simply to press a foot pedal when the probe stimulus appeared, alternating feet over successive trials. The participants were completely unaware of the meaning of the stimuli.

Verbal condition. During the learning phase, prior to scanning, the participants received two verbal descriptions of figures, one labeled "Figure 1" and the other labeled "Figure 2." The descriptions were presented on separate sheets of paper that were mounted on the screen, which the participants were allowed to remove in order to hold each description directly in front of them if they desired. They were asked to study these sheets until they could recite each description verbatim when given the corresponding number label. The participants had to recall each description perfectly twice in succession before going on to the next part of the study phase. Once they had memorized the descriptions, they read instructions on the computer's screen that specified how the figures should be visualized on the basis of the descriptions. Next, the participants were shown the donut grid and were asked to visualize figures that corresponded to the descriptions they had just memorized. After the participants had claimed that they had visualized the patterns successfully, they were told to keep the figures in mind and went on to the response training session.

In the response training session, the participants were asked simply to press the corresponding foot pedal when the word *yes* or *no* was presented briefly (for 100 msec) on the screen; they were told which pedal corresponded to each response. After each word had appeared, the screen remained blank until the participants responded, and the next stimulus was presented 1 sec later. Thirty-one stimuli were presented in this session, and the computer beeped each time the participants answered incorrectly.

Following response training, the trials began. Each trial consisted of the same series of events that was presented in the baseline condition. In this task, however, the participants were asked to visualize the figure (within the brackets) labeled by the number, and then to press the *yes* pedal if the X-probe would have fallen on the appropriate figure, had it been in the brackets, or the *no* pedal if the X-probe would not have fallen on the figure. A new trial was presented immediately after each response.

The participants took part in four practice trials prior to scanning, to learn the task properly; the computer beeped if they made an error during practice but did not provide such feedback during the actual test trials. If the participants made any errors during practice, the task was again described. However, to ensure that the par-

ticipants encoded the figures solely on the basis of the verbal description, these descriptions were presented only at the beginning of the learning phase, before the participants knew how to construct the figures in the donut grid.

At the end of the scanning session, after the participants had performed all the conditions and had left the scanner, they were asked to draw the visualized figures, to ensure that they had been learned correctly. If one or both of these figures were drawn incorrectly, we assumed that the figures had not been visualized properly and discarded that participant's data. Similarly, if the participants responded incorrectly on 40% or more of the trials, we had to assume that they had not followed the instructions properly, had failed to learn the figures, or had difficulties creating an image; hence, we discarded their data.

Amalgamation condition. In the learning phase, the participants studied one segment of one figure (each figure was labeled throughout the study phase as either "Figure 1" or "Figure 2") that was positioned within a set of four brackets; when they had memorized the length and position of that segment, they pressed a foot pedal, and 1 sec later, that segment disappeared and another segment was presented; the second segment would have been connected to the first if both had been visible. This procedure continued until the entire set of segments for a figure had been viewed. The participants were able to study each segment for as long as they wished; however, they never saw the entire figure at once but saw it only one segment at a time. The participants were instructed to mentally "glue" the segments together to form a mental image of the complete figure. After seeing all of the segments for a given figure, the participants were given a blank set of brackets with the number 1 or 2 below it, and were to visualize the corresponding completed pattern. The learning procedure was repeated three times, so the participants could correctly visualize and adjust their image of each figure.

Once the participants had reported accurate images of each figure, they were asked to keep these images in mind and to perform a response key training session, which was identical to the response key training session in the verbal condition. Immediately following response training, the participant received four practice trials (with feedback for accuracy, as in the verbal condition). Following this, the test trials began. These trials were identical to those in the verbal condition; however, in this case, the participants were asked to use the corresponding mental images of the two patterns they mentally "glued" together. Once again, after the participants had completed all the conditions and had left the scanner bed, they were asked to draw the figures they had created in their mind's eye. If the participants drew one or more of the figures incorrectly or had responded incorrectly on 40% or more of the trials, we assumed that they had been unable to visualize the figures or follow the instructions properly and, therefore, had to discard their data.

PET Procedure

The participants were tested in a GE Scanditronix PC4096 15-slice whole-body tomograph, which was operating in the stationary mode (see Rota Kops, Herzog, Schmid, Holte, & Feinendegen, 1990). A certified radiology technologist positioned each participant into the PET scanner and fitted him with a thermoplastic face mask (TruScan, Annapolis, MD), which, as was mentioned above, was designed to minimize head movements during the scanning procedure. A fixed vertical and horizontal laser system helped us to position the participant correctly, measuring head location relative to the cantho-meatal line. Each participant was also fitted with nasal cannulae and a vacuum mask; the cannulae delivered the $^{15}\text{O}-\text{CO}_2$, and the mask, attached to a vacuum system, removed any excess radioactivity. Each emission scan lasted 1 min and a 10-min interval separated the conditions, allowing the $^{15}\text{O}-\text{CO}_2$ to decay. The tomograph was located in a private room used specifically for cognitive neuroscience testing. The tomograph produced 15 con-

secutive slices that were 6.5 mm apart, with axial resolution at 6.0 mm full width at half maximum (FWHM). Before emission scans took place, three transmission scans were performed with an orbiting rod source over the axial extent covering the cerebral hemispheres and the cerebellum. Twenty measurements were made on each scan. The procedure for each scan involved (1) starting the camera acquisition program (which measured residual background from the previous study), (2) starting stimulus presentation 15 sec later, (3) administering the ^{15}O - CO_2 gas 15 sec after stimulus presentation had begun, and (4) ceasing gas administration and scanning 60 sec later. The concentration of the delivered gas was 2,800 MBq/l at a flow rate of 2 l per minute and diluted by mixture with room air, so that the measured peak count rate from the brain was 100,000–200,000 per second. Images of relative blood flow were based on Scans 4–16, which were summed after reconstruction. For additional details on the scanning procedure, see Kosslyn et al. (1994).

PET Statistical Analysis

Statistical analysis followed the theory of statistical parametric mapping (Friston, Frith, Liddle, & Frackowiak, 1991; Friston et al., 1995; Worsley, Evans, Marrett, & Neelin, 1992). Data were analyzed with the SPM99 software package (from the Wellcome Department of Cognitive Neurology, London). As a first step, the images from each condition were summed over the entire scanning period for that condition and then were resliced to a horizontal plane. All the scans were then automatically realigned to a reference point, which was based on the initial position of that participant during the first scan. This least-squares procedure corrected for head tilt within the scanner gantry, as well as for small head movements across the entire scanning session. The PET data at each voxel then were normalized by the global mean and fit to a linear statistical model by the method of least squares. The gray-matter threshold was set to the default of 0.8. The analysis considered cognitive state (i.e., scan condition) as the main effect and participants as a block effect. Hypothesis testing was performed using the method of planned contrasts at each voxel. This method fits a linear statistical model, voxel by voxel, to the data; hypotheses are tested as contrasts in which linear compounds of the model parameters (e.g., differences) are evaluated using t statistics. Data from all three conditions were used to compute the contrast error term. The resulting images were then spatially smoothed with a two-dimensional Gaussian filter (15 mm wide, FWHM), thus removing high-frequency noise. A z threshold of 3.09 (corresponding to $p = .001$) was considered statistically significant, provided that five contiguous voxels reached this criterion. The combined threshold was thus well below the .001 range.

RESULTS

We began by analyzing the response times and error rates in order to provide evidence that the participants did, in fact, perform the tasks, which would give us confidence that the PET results reflected the kinds of processing we wanted to study.

Behavioral Results

Response times and error rates were assessed in separate analyses of variance (ANOVAs). The analyses included condition (verbal or amalgamation), and probe position (early or late in the sequence in which the segments had been described or presented) as factors. The response times were trimmed so that any time greater than 2.5 times the participant's mean in that condition was excluded. In addition, trials in which the response

time was less than 200 msec were excluded, because such times could not have reflected the processing of interest. These procedures led us to trim an average of 3.1% of the trials in the verbal condition and 1.5% of the trials in the amalgamation condition.

Response times. We expected that the participants would generate the images a segment at a time and, thus, evaluate probes positioned on or near segments visualized early in the sequence more quickly than those positioned on or near later segments. And in fact, the participants required an average of 1,256 msec for early probes versus 1,423 msec for late probes [$F(1,15) = 5.50, p < .05$]. These results mirror those of Kosslyn, Cave, Provost, and von Gierke (1988; Kosslyn, Maljkovic, et al., 1995) and provide evidence that the participants did, in fact, generate images in these tasks. No other main effect or interaction was significant. The two conditions required comparable amounts of time in general, with means of 1,410 msec for the amalgamation condition and 1,268 msec for the verbal condition ($p > .1$).

Error rates. The participants made errors on 14.6% of the trials in the amalgamation condition versus 9% of the trials in the verbal condition, although neither this difference nor any other main effect or interaction was significant ($p > .05$ in all cases).

PET Results

We not only compared the imagery conditions with each other, but also compared each condition with the baseline.

Comparing imagery conditions. We began with the most straightforward method of determining whether visual mental images can be formed in two ways. In this analysis, we directly contrasted regional cerebral blood flow (rCBF) in the verbal condition with that in the amalgamation condition and vice versa.

Figure 2 presents the locations and coordinates of areas that included at least five contiguous voxels with a z value over 3.09 when the conditions were directly contrasted. Coordinates were identified in the Montreal Neurological Institute (MNI) system. These results are easy to summarize. First, when we subtracted rCBF in the amalgamation condition from that in the verbal condition, we found greater rCBF in two regions, both of which were in the frontal lobes. Moreover, both of these regions are in the LH, which is consistent with the view that categorical spatial relations were processed. Second, when we performed the reverse contrast, we found that only two areas were activated more strongly in the amalgamation condition than in the verbal condition, one of which was in the frontal lobes and the other of which was in the inferior parietal lobule. Both of these regions were in the RH, which is consistent with the view that coordinate spatial relations were processed. (A third RH point at coordinates 22 22 18 appeared to be on a white matter tract and, thus, was not considered in this analysis.)

Although these results are consistent with our lateralization hypothesis, these direct contrasts do not confirm

MNI COORDINATES, NUMBER OF VOXELS, AND z SCORES FOR FOCI OF ACTIVATION ILLUSTRATED ABOVE:

- ▲ Dorsolateral Prefrontal Cortex
Inferior Frontal (BA 45/46): MNI coord. $[-44\ 40\ 18]$; Voxels [7]; z score [3.18]
- ▲ Insula/Inferior Frontal (BA 47): MNI coord. $[-34\ 18\ -4]$; Voxels [75]; z score [3.95]
- Inferior Parietal (BA 40): MNI coord. $[56\ -28\ 46]$; Voxels [16]; z score [3.38]
- Insula/Inferior Frontal (BA 44/45): MNI coord. $[46\ 6\ 14]$; Voxels [56]; z score [3.62]

Figure 2. Points of maximal activation in, and MNI coordinates of, areas that were activated more in the amalgamation condition than in the verbal condition and vice versa. Each tick mark represents 20 mm. MNI coordinates, the number of voxels associated with the activation, and the z score of the activation are also provided for all foci.

that categorical processing takes place preferentially in the LH and coordinate processing in the RH. They only show that there is significantly greater activation in one hemisphere for one condition than in the corresponding locus in the same hemisphere *for the other condition*. However, it is possible that this difference between the two conditions (in one location of one hemisphere) does not identify activation that is truly lateralized; it is possible that activation in these locations is not significantly different from that in the corresponding location in the contralateral hemisphere. The strongest (and most stringent) evidence for hemisphere differences between the tasks would be a condition $\times$ hemisphere interaction, which would reveal locations that are activated differently in the two conditions in the two hemispheres. Thus, in the next analysis, rather than simply documenting main effects, we asked whether condition $\times$ hemisphere interactions can be observed for individual loci—a question rarely asked in neuroimaging research. Specifically, we wished to discover whether any individual loci are relatively more active in the LH during the verbal condition and in the RH during the amalgamation condition. By showing such a difference in the same area, we can rule out any possible artifacts due to overall levels of activation during the conditions or inappropriate comparisons. We need to show not only that a specific region has greater activation in the amalgamation condition than in the verbal condition, for example, but also that this difference is greater than the difference in the con-

tralateral hemisphere. Otherwise, there's no true hemispheric effect. In other words, we need to document not simply that rCBF differed in the two hemispheres, but also that this difference was modulated by condition.

In order to carry out this analysis, we first flipped the images along the x -axis in order to reverse the left and the right hemispheres of the brain. We then set up the statistical contrasts, so that both the conditions and the hemispheres were contrasted. Thus, we applied the contrast *amalgamation nonflipped minus amalgamation flipped minus verbal nonflipped minus verbal flipped*. In accordance with the recommendations of Friston (SPM archives, May 18, 2000), we used a multigroup, different-conditions model within SPM to best estimate the hemispheric differences. This contrast resulted in a statistical map for the condition $\times$ hemisphere interaction. Changing all the valences of the contrast from positive to negative and vice versa necessarily results in the opposite contrast, *verbal nonflipped minus verbal flipped minus amalgamation nonflipped minus amalgamation flipped*, with only the laterality of the significant regions changed in the results.

The findings from this interaction contrast were straightforward. We identified a region on the border of the inferior frontal lobe and the precentral gyrus—Brodmann area (BA) 44/6—located at MNI coordinates $(\pm) 46\ 6\ 16$ in which rCBF was greater in the LH than in the RH in the verbal condition and vice versa in the amalgamation condition. The z score for the contrast pair was 3.75 ($p < .001$).

Because statistically significant interactions with two or more variables are difficult to obtain with a limited sample size, we also will report here (see Table 1) the areas of activation obtained with a more liberal threshold of 2.56 (corresponding to a p value of .01, uncorrected). As we have noted in previous articles in which we have reported such results (see, e.g., Thompson, Kosslyn, Sukel, & Alpert, 2001), these regions should be interpreted with caution, but we provide them here for researchers who may have specific hypotheses with respect to these individual areas. At the z threshold of 2.56, approximately 97% of the total number of activated voxels are lateralized—and this lateralization is, without question, in the predicted direction.

Verbal versus baseline. We next compared rCBF in the verbal condition with that in the baseline condition. These results are presented in Table 2 and Figure 3 (top panel). We found a large region of activation, which extended from the frontal to the superior/middle temporal and posterior parietal and occipital lobes. We expected and found centroids of activation in the ventral system, where representations of object shapes apparently are stored; the fusiform gyrus was activated bilaterally, and the left occipito-temporal junction was activated. However, it is important to note that we did *not* find foci of activation in either Broca's or Wernicke's area—despite the fact that, in this condition, the participants memorized verbal descriptions of the shapes.

In addition, some aspects of these results were similar to those found by Kosslyn et al. (1998), where a categorical perceptual encoding task was used. In this task, participants decided whether an X was *above* or *below* a bar (the stimuli were always presented on the screen and viewed for as long as the participant needed to make the judgment [Experiment 1] or were presented very briefly on the screen [Experiment 2]). Specifically, we found activation in the middle frontal cortex, in the occipito-temporal junction, and covering much of the parietal lobes, near the occipito-parietal junction. We expected activation of structures involved in attention, which would

be required to monitor what one is creating as the image is built up from the description. Activation in the occipito-parietal junction (near the precuneus and the posterior cingulate) may play a role in attention and attention shifting (see Mesulam, 1981; Posner & Petersen, 1990), and activation in the vicinity of these areas has been reported in other imagery experiments (e.g., Mellet et al., 1996; Mellet, Tzourio, Denis, & Mazoyer, 1995).

Amalgamation versus baseline. We next compared rCBF in the amalgamation condition with that in the baseline condition. These results are presented in Table 3 and Figure 3 (bottom panel). We detected a large swath of bilateral activation spreading from the middle and inferior frontal lobes back through middle and superior portions of the temporal cortex to the posterior parietal and occipital lobes. The maximal point of activation was located in the inferior portion of the right frontal lobe, in the vicinity of the caudate nucleus. Previous findings (Kosslyn, Thompson, & Alpert, 1997) led us to expect activation in the occipito-temporal junction, and we found bilateral fusiform gyrus activation, on the border of BAs 37 and 19, which was clearly part of the ventral system (which processes object properties, such as shape and color, and stores visual memories; see, e.g., Haxby et al., 1991; Levine, 1982; Sergent, Ohta, & MacDonald, 1992; Sergent, Zuck, Levesque, & MacDonald, 1992; Tanaka, Saito, Fukada, & Moriya, 1991; Ungerleider & Mishkin, 1982; see Grill-Spector, 2003, for a review). In addition, Area 18 activation was included in the large cluster noted above; according to our theory (Kosslyn & Thompson, 2003), when visual memories are reactivated, the details of the shape of an object are made explicit in the early visual cortex (corresponding to Areas 17 and 18). Moreover, we found strong activation in the occipito-parietal junction. Activation in this area is probably related to the allocation of spatial attention. In addition, a centroid of activation was detected in the left inferior parietal lobule, which was also reported by Kosslyn et al. (1998), who used a spatial relations perceptual task (compared with a baseline task that was similar to the one used in this

Table 1
Locations and Numbers of Voxels for Clusters of Five or More Contiguous Voxels That Had z Values Over 2.56 for a Contrast Analysis Examining the Condition $\times$ Hemisphere Interaction

Region (BA)	Voxels	MNI Coordinates			z Score
		x	y	z	
Inferior frontal/precentral gyrus (44/6)	100*/364**	46	6	16	3.75
Postcentral gyrus (3-2-1)	91	60	-18	36	2.88
Inferior/middle temporal (20/21)	75	52	-20	-6	2.79
Cingulate gyrus (24/31)	41	16	-24	44	2.77
Cuneus (18/19)	20	-24	-82	34	2.73

Note—The x coordinates vary, depending on the valence of the contrast weights; for simplicity, we present here only the results from one side of the contrast, specifying positive weights associated with more activation in the amalgamation condition. Note that the region at MNI coordinates 46 6 16 was the only one to attain statistical significance according to our threshold z score of 3.09. Other regions (with z scores of 2.56–3.09) are presented for researchers who have specific hypotheses about these regions. The regions of activation were estimated on the basis of transformation to Talairach coordinates (Talairach & Tournoux, 1988) and on an overlay of the PET emission data to a standardized MNI template in SPM. Brodmann's areas (BAs) are provided in parentheses when available. *Threshold at 3.09. **Threshold at 2.56.

Table 2
Locations and Numbers of Voxels for Clusters of Five or More Contiguous Voxels
That Had z Values Over 3.09 When Activation in the Baseline Condition was Subtracted
From That in the Verbal Condition

Region	Voxels	MNI Coordinates			z Score
		x	y	z	
Inferior/middle frontal/insula*	35597*	-38	20	16	6.76
Occipito-parietal junction*		24	-72	40	6.59
Area 18*		-26	-78	6	6.37
Occipito-temporal junction/fusiform gyrus (BA 37)	140	-46	-50	-4	3.82
Superior temporal (BA 38)	10	-56	18	-16	3.49
Fusiform gyrus (BA 37)	19	36	-40	-6	3.38
Fusiform gyrus (BA 37)	14	-18	-40	-6	3.27

Note—Coordinates are presented according to the Montreal Neurological Institute (MNI) template. Brodmann's areas (BAs) are provided in parentheses when available. The regions of activation were estimated on the basis of transformation to Talairach coordinates (Talairach & Tournoux, 1988) and on an overlay of the PET emission data to a standardized MNI template in SPM. *Part of a larger region encompassing a swath of activation from the frontal lobes through the superior portion of the temporal lobe to the posterior parietal and occipital cortices.

study). Unlike Kosslyn et al. (1998), we did not find a separate centroid of activation associated with the right posterior parietal cortex; however, the activation centered around the right occipito-parietal junction was extensive, spreading far into the posterior parietal lobes. Many activated regions within the parietal cortex may have been encompassed by this large spread of activation.

We also found activation in the right dorsolateral prefrontal cortex (DLPFC). This region apparently plays a role in actively accessing stored information (see Kosslyn, Thompson, & Alpert, 1995; Tulving, Kapur, Craik, Moscovitch, & Houle, 1994). Such processing would be required to access stored locations of the segments when one generates an image. In addition, we detected activation in the parahippocampal gyrus, which apparently plays a role in spatial memory (Ploner et al., 2000). We also found points of activation in the medial frontal cortex (in the vicinity of the anterior cingulate) and in the precentral cortex.

The two panels of Figure 3 look so similar that we worried we had made an error in the analyses (after checking—several times!—we are confident that these results are correct). The two conditions clearly rely largely on the same underlying mechanisms, essentially evoking a visual imagery network. These similar patterns of activation in the two conditions suggest that many of the same processes are used to form an image based on a verbal description and an image based on stored visual information. To estimate the amount of overlap in activation between the conditions, we grouped the results from both conditions and compared them with the baseline. This analysis allowed us to discover in which areas activation was present either in both conditions or in only one of the two conditions. With a threshold of 3.09, we found a total of 41,468 activated voxels. When we compared this number with the number of voxels activated in each of the conditions separately versus the baseline, as shown in Tables 2 and 3, we found that 35,597 voxels were activated in the verbal condition and 31,015 voxels were activated in the amalgamation condition. We then

compared the number of voxels that were significantly different from baseline in each condition separately with the number that were activated when the results from both conditions were pooled. This calculation led to an estimated overlap of between 74.8% (amalgamation condition) and 85.8% (verbal condition). Using a simpler approach, we examined the number of clusters that were found in both the verbal and the amalgamation conditions versus the baseline and the number that were found when the conditions were directly contrasted. We found only four clusters that were different, and found 17 that were activated when we compared the pooled experimental conditions with the baseline. This simple procedure leads us to estimate an overlap of 81%. This finding is consistent not only with the calculations above, but also with findings reported by Denis and his colleagues (e.g., Denis & Cocude, 1992), which show that well-structured verbal descriptions can allow one to produce accurate visual images.

In addition, we examined whether the conditions engendered specific and similar patterns of activation by raising the z score threshold for both conditions, as compared with the baseline. If the two test conditions do, in fact, activate a similar network of regions, we would expect the same regions to be activated most strongly. In fact, when we raised the z score threshold from 3.09 to 5.0, we found a very similar pattern of activation in the two test conditions, as compared with the baseline. With this higher threshold, a similar number of voxels was activated in both conditions (approximately 13,000), and the statistical maps revealed a pattern of activation that extended bilaterally from the parietal lobes to the middle and inferior portions of the frontal lobes.

DISCUSSION

The purpose of this study was to investigate whether there are (at least) two distinct ways in which visual mental images can be generated. The logic of this study hinged on the idea that if different methods are used to generate

Figure 3. Results from the comparison of images formed in the verbal condition (top panel), which were learned as descriptions, and those formed in the amalgamation condition (bottom panel), which were learned as sets of isolated segments that were mentally “glued together,” with the baseline condition. The “glass brain” diagrams are adapted from SPM99. Sagittal (top left), coronal (top right), and axial (bottom) slices of the brain are visible in each panel. These diagrams clearly show the spread of activation from the frontal lobes to the posterior parts of the parietal and occipital lobes. Tables 2 and 3 provide the Montreal Neurological Institute coordinates of the points of maximal activation.

images following different types of learning, at least partially different brain systems should be activated. And in fact, we found differences in activation between the two test conditions, with relatively more activation in the RH when images were formed after stimuli had been learned by mentally “gluing” segments together than when they were formed after verbal descriptions were memorized, and vice versa in the LH. This finding that different mental strategies can invoke different brain processes, even

when the task itself is identical, is consistent with previous findings. For example, different brain areas were activated after participants were trained, implicitly (Wraga, Thompson, Alpert, & Kosslyn, 2003) or explicitly (Kosslyn, Thompson, Wraga, & Alpert, 2001), to mentally rotate objects as if they were either rotating them with their hands or observing them being rotated by an external force.

It is useful to consider the present findings in the context of those reported by Kosslyn et al. (1998), who examined the neural bases of perceptually encoding categorical and coordinate spatial relations. Specifically, participants viewed stimuli that contained a bar and an X mark, and judged whether the X was above or below the bar (the categorical condition) or within 0.5 in. of it (the coordinate condition). Kosslyn et al. (1998, Experiments 1 and 2) reported no greater parietal activation during the categorical task than during the coordinate task. In the present study, we found greater parietal activation during the amalgamation (coordinate) condition when it was directly compared with the verbal (categorical), but no greater parietal activation with the reverse comparison. Because the stimuli were encoded prior to scanning, it is possible that the major work of the parietal lobes in the categorical task had already been accomplished before the scan session. Alternatively, left parietal activation in the amalgamation condition may have masked inferior parietal activation in the verbal condition. Given the large spread of activation in both conditions from the peak point in the frontal lobes bilaterally to the temporal and posterior parietal and occipital lobes, this explanation is plausible. In addition, Kosslyn et al. (1998) reported bilateral parietal activation in the coordinate task used in that study. Jager and Postma (2003) noted that studies performed to date support the general view of a left/right specialization for categorical and coordinate spatial relations processing, respectively, and that the parietal lobes appear to be involved; however, they did not find this evidence completely convincing, particularly for categorical processing, and our results are consistent with this conclusion.

We were surprised by some of the specific sites that were (or were not) activated differently in the two test conditions, as compared with baseline. First, we found independent activation of DLPFC only in the amalgamation condition, and this region is within the area that Jonides and Smith (1994) and Smith and Jonides (1999) identified with a component of a *working memory* system. It is not clear why the verbal condition failed to engender a clear centroid of activation in this structure, particularly in the LH, given past findings (e.g., Kosslyn et al., 1998) in which left-DLPFC activation was clearly observed with verbal tasks; again, we note that there was a large spread of activation bilaterally in the frontal lobes when the verbal condition was compared with the baseline. The maximal point of activation was not far from the DLPFC. When the data from the amalgamation condition were directly contrasted with those from the verbal condition, we also found evidence of activation in a

Table 3
Locations and Numbers of Voxels for Clusters of Five or More Contiguous Voxels That Had
***z* Values Over 3.09 When Activation in the Baseline Condition Was Subtracted From That in the**
Amalgamation Condition

Region	Voxels	MNI Coordinates			<i>z</i> Score
		<i>x</i>	<i>y</i>	<i>z</i>	
Inferior frontal/caudate*	31015*	24	16	22	6.98
Area 18*		-26	-78	8	6.70
Occipito-parietal junction*		24	-72	40	6.69
Inferior parietal lobule (BA 40)	40	-48	-42	62	4.19
Anterior cingulate/medial frontal (BA 32/8)	90	-8	28	38	4.13
Parahippocampal gyrus (BA 36)	69	-20	-38	-6	3.77
Fusiform gyrus (BA 37/19)	20	36	-42	-4	3.76
Dorsolateral prefrontal cortex (BA 46/10)	20	50	54	6	3.67
Precentral gyrus (BA 4/6)	7	70	2	20	3.35
Fusiform gyrus (BA 37)	5	-40	-42	-2	3.26

Note—Coordinates are presented according to the Montreal Neurological Institute (MNI) template. Brodmann's areas (BAs) are provided in parentheses when available. The regions of activation were estimated based on transformation to Talairach coordinates (Talairach & Tournoux, 1988) and by overlaying the PET emission data to a standardized MNI template in SPM. *Part of a larger region encompassing a swath of activation from the frontal lobes through the superior portion of the temporal lobe to the posterior parietal and occipital cortices.

region on the border of the insula and the inferior frontal cortex, corresponding to BAs 44 and 45. In addition, we found activation in the other areas that Jonides and Smith identified as the working memory circuit. Second, it is of note that we found relatively greater activity in left Area 47 during the verbal condition, which has been associated with linguistic semantic processing (McDermott, Petersen, Watson, & Ojemann, 2003), as well as with spatial working memory (Smith & Jonides, 1998). Although words *per se* probably were not stored, their meanings probably were. We note that brain areas typically associated with reading *per se*—specifically, in the left posterior superior temporal (BAs 21/22/42), the cerebellum (Fiez & Petersen, 1995; Turkeltaub, Eden, Jones, & Zeffiro, 2002), and the inferior prefrontal cortices (Marinkovic, 2004)—were not activated in the verbal condition.

Moreover, the single area that clearly reversed hemisphere activation during the two conditions was located in the inferior portion of the frontal lobe, near the precentral gyrus. This finding is consistent with that of Kosslyn et al. (1998), who found that left frontal regions were associated with categorical spatial processing and right frontal regions were associated with coordinate spatial processing. Kosslyn et al. (1998) found that the clearest hemispheric difference between the two conditions was in the frontal, rather than the parietal, lobes. This result is also consistent with a finding reported by Jonides et al. (1993). In one task, they asked participants to study three dots on a screen and remember their locations. Later, the participants were to judge whether a circle would have covered one of the dots. In order to perform the task, the precise metric locations of the dots would have to be stored, and thus, coordinate spatial processing would play a role in this task. Thus, it is of interest that this task engendered activation in the right premotor area (Area 6), as well as in right inferior portions of the frontal lobe (Area 47). The region we found

here is located between the centers of Areas 47 and 6 and may be involved in spatial working memory (see Smith & Jonides, 1998).

Although we did find evidence that distinct processing underlies the two ways of generating images, the differences between the conditions were subtle. More striking was the evidence for many common processes. Specifically, in both conditions, we detected a nearly identical swath of bilateral activation spreading from the middle and inferior frontal lobes back through the middle and superior portions of the temporal cortex to the posterior parietal lobe. This region encompasses many areas that previously have been identified with visual mental imagery. According to our theory (see, e.g., Kosslyn, 1994), visual mental images arise when a pattern of activation is induced in topographically organized regions of the occipital lobe; such patterns of activation depict the geometric properties of the shape of an object. Images are created on the basis of activation propagated backward from the inferior temporal lobes, where visual memories are stored in a compressed form. In addition, posterior portions of the parietal lobes are invoked when spatial properties of an object must also be processed—such as when parts of an object are arranged to form a whole, as in the present experiment. Moreover, according to this theory, the frontal lobes and, particularly, the DLPFC are involved in accessing stored information, such as the locations of segments.

The fact that the two conditions engendered very similar overall patterns of activation allows us to address another issue. The logic of our investigation rests on the assumption that mental images are representations like those evoked during the early phases of perception, but created on the basis of information stored in memory, instead of on the basis of input from a sensory organ (as occurs during perception); such representations give rise to the experience of perceiving in the absence of the ap-

propriate sensory input (e.g., *seeing with the mind's eye* or *hearing with the mind's ear*). A representation, in this sense, is a physical state that serves as a repository of information (and occurs within a system of processes that can interpret and use such representations; see, e.g., Kosslyn & Koenig, 1995). This *representational view* of imagery has recently been challenged by *perceptual activity theory*, largely developed by O'Regan (1992) and his colleagues. As Thomas (2002) puts it, "Visual experience, O'Regan holds, arises not from the presence of representations in the brain but from the active exercise of our 'mastery of the relevant sensorimotor contingencies' (O'Regan & Noë, 2001) as we explore our visual surroundings. Perceptual activity theory holds that imagery arises from vicarious exercise of such mastery; a sort of play-acting of perceptual exploration (Thomas, 1999)" (p. 211).

However, even such "play-acting" must arise from activity in the brain. And crucially, different sorts of play-acting should be supported by different patterns of brain activation. In our tasks, the brain activity underlying the "vicarious exercise" of encoding verbal descriptions versus segments of shapes would not involve encoding categorical and coordinate representations. Rather, we should have found activation in Broca's and Wernicke's areas when people were asked to recall the shapes encoded verbally, which we did not. Moreover, we should have found activation in the frontal eye fields when people were asked to recall the locations of individual segments in the amalgamation condition, which we did not; according to the play-acting view, they should have moved their eyes as they would have in perception when carrying out the same activity (Brandt & Stark, 1997; Laeng & Teodorescu, 2002), but there was no evidence of such processing, which apparently is not a necessary aspect of image generation. Indeed, given the gross differences between the processes that would be used in "perceptual exploration" of the two sorts of stimuli, perceptual activity theory would have led us to expect large differences between the two conditions, but we found only subtle differences in rCBF between them. The results rather suggest that the participants were able to generate the same image representation in the two conditions and could use it in the same ways once it was generated. In short, if the same sorts of representations are produced by both methods, we would expect overall similar patterns of activation, but to the extent that there are differences, they would reflect differences in the mental functions used to produce the images. It is not the case that any combination of similarity and difference in activation would confirm our predictions.

Could the similarity of activation in the two conditions simply reflect massive activation in general? If so, the similarity would simply be an artifact of overall high levels of arousal or the like. In fact, many commonly activated areas—such as Broca's and Wernicke's areas—were not activated in either condition. Moreover, the observed pattern of similarity was not random or haphazard. First, the

pattern of activation makes sense within the general theory of mental imagery outlined by Kosslyn (1994). Second, when we raised the z score threshold to 5.0, the remarkable similarity of the pattern was preserved—despite the fact that relatively few voxels were deemed to be "activated." Finally, the fact that many z scores were actually above 5.0 (ranging up to 6.76) demonstrates that this pattern not only is similar across the test conditions, but also is very robust. That said, to demonstrate definitively that the two patterns of activation are similar because of the specific underlying processes they draw upon, we probably would need to include a third test condition, perhaps involving verbal processing, and show that this condition elicited dramatically different patterns of rCBF for the same participants.

The selectivity of the present findings is also supported by the fact that our results are largely consistent with those reported by Mellet et al. (2002), who found a very similar pattern of activation when participants scanned an imaged map that they had generated from verbal descriptions and when they scanned a similar mental map that they had learned from visual inspection. Nevertheless, some differences emerged between their two conditions: In the group that had learned the map on the basis of a verbal description, the angular gyrus and Broca's and Wernicke's areas were activated, whereas in the group that had learned the map on the basis of visual inspection, the right medial temporal lobe was activated. The researchers suggested that the scanning task engaged the same network for both groups but that traces of the modality of encoding remained at the time that the task was performed. Why did Mellet et al. (2002) find activation of traditional language areas and we did not? Whereas our stimuli could be visualized from the verbal descriptions quickly (typically within 2 or 3 min) and easily retained, the participants in Mellet et al.'s (2002) study spent up to 20 min (10–13 min on average) encoding their complex descriptions of landscapes. The participants may have remembered verbal encodings of the locations and used them as self-generated prompts during scanning. Note that this use of verbal information is a far cry from the *mimicking encoding* idea espoused by perceptual activity theory.

In short, three general conclusions can be drawn from the present findings. First, visual mental images can be formed using at least two methods; this conclusion is compatible with results in the current literature on the processing of categorical and coordinate spatial relations. Second, despite the different sorts of perceptual exploration that would be guided by the stored information, remarkably similar overall brain activation occurred in both conditions. This finding is consistent with the view that imagery is not simply the act of mimicking perceptual exploration. Third, the systems underlying visual mental imagery are likely to be as complicated as the systems underlying visual perception—which is not surprising, because imagery and perception draw on many of the same mechanisms.

REFERENCES

- BRANDT, S. A., & STARK, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, **9**, 27-38.
- CHABRIS, C. F., & KOSSLYN, S. M. (1998). How do the cerebral hemispheres contribute to encoding spatial relations? *Current Directions in Psychological Science*, **7**, 8-14.
- COSTIN, D. (1988). MacLab: A Macintosh system for psychology labs. *Behavior Research Methods, Instruments, & Computers*, **20**, 197-200.
- DENIS, M., & COUDE, M. (1992). Structural properties of visual images constructed from poorly or well-structured verbal descriptions. *Memory & Cognition*, **20**, 497-506.
- FARAH, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition*, **18**, 245-272.
- FARAH, M. J. (1986). The laterality of mental image generation: A test with normal subjects. *Neuropsychologia*, **24**, 541-551.
- FIEZ, J. A., & PETERSEN, S. E. (1995). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences*, **95**, 914-921.
- FRISTON, K. J., FRITH, C. D., LIDDLE, P. F., & FRACKOWIAK, R. S. J. (1991). Comparing functional (PET) images: The assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism*, **11**, 690-699.
- FRISTON, K. J., HOLMES, A. P., WORSLEY, K. J., POLINE, J.-P., FRITH, C. D., & FRACKOWIAK, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, **2**, 189-210.
- GRILL-SPECTOR, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, **13**, 159-166.
- 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. *Proceedings of the National Academy of Sciences*, **88**, 1621-1625.
- HELLIGE, J. B., & MICHIMATA, C. (1989). Categorization versus distance: Hemispheric differences for processing spatial information. *Memory & Cognition*, **17**, 770-776.
- JAGER, G., & POSTMA, A. (2003). On the hemispheric specialization of categorical and coordinate spatial relations: A review of the current evidence. *Neuropsychologia*, **41**, 504-515.
- JONIDES, J., & SMITH, E. E. (1994). Working memory in humans: Neuropsychological evidence. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1009-1020). Cambridge, MA: MIT Press.
- JONIDES, J., SMITH, E. E., KOEPE, R. A., AWH, E., MINOSHIMA, S., & MINTUN, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature*, **363**, 623-625.
- KOSSLYN, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, **94**, 148-175.
- KOSSLYN, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- 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., BRUNN, J. L., CAVE, K. R., & WALLACH, R. W. (1984). Individual differences in visual imagery: A computational analysis. *Cognition*, **18**, 195-243.
- KOSSLYN, S. M., CAVE, C. B., PROVOST, D. A., & VON GIERKE, S. M. (1988). Sequential processes in image generation. *Cognitive Psychology*, **20**, 319-343.
- KOSSLYN, S. M., CHABRIS, C. F., MARSOLEK, C. M., & KOENIG, O. (1992). Categorical versus coordinate spatial representations: Computational analyses and computer simulations. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 562-577.
- KOSSLYN, S. M., HOLTZMAN, J. D., FARAH, M. J., & GAZZANIGA, M. S. (1985). A computational analysis of mental image generation: Evidence from functional dissociations in split-brain patients. *Journal of Experimental Psychology: General*, **114**, 311-341.
- KOSSLYN, S. M., & KOENIG, O. (1995). *Wet mind: The new cognitive neuroscience* (2nd ed). New York: Free Press.
- KOSSLYN, S. M., KOENIG, O., BARRETT, A., CAVE, C. B., TANG, J., & GABRIELI, J. D. E. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 723-735.
- KOSSLYN, S. M., MALJKOVIC, V., HAMILTON, S. E., HORWITZ, G., & THOMPSON, W. L. (1995). Two types of image generation: Evidence for left and right hemisphere processes. *Neuropsychologia*, **33**, 1485-1510.
- KOSSLYN, S. M., & THOMPSON, W. L. (2003). When is early visual cortex activated by visual mental imagery? *Psychological Bulletin*, **129**, 723-746.
- KOSSLYN, S. M., THOMPSON, W. L., & ALPERT, N. M. (1995). Identifying objects at different levels of hierarchy: A positron emission tomography study. *Human Brain Mapping*, **3**, 107-132.
- 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., THOMPSON, W. L., GITELMAN, D. R., & ALPERT, N. M. (1998). Neural systems that encode categorical versus coordinate spatial relations: PET investigations. *Psychobiology*, **26**, 333-347.
- KOSSLYN, S. M., THOMPSON, W. L., SHEPARD, J. M., GANIS, G., BELL, D., DANOVITCH, J., WITTENBERG, L. A., & ALPERT, N. M. (2004). Brain rCBF and performance in visual imagery tasks: Common and distinct processes. *European Journal of Cognitive Psychology*, **16**, 696-716.
- KOSSLYN, S. M., THOMPSON, W. L., WRAGA, M., & ALPERT, N. M. (2001). Imagining rotation by endogenous versus exogenous forces: Distinct neural mechanisms. *NeuroReport*, **12**, 2519-2525.
- LAENG, B. (1994). Lateralization of categorical and coordinate spatial functions: A study of unilateral stroke patients. *Journal of Cognitive Neuroscience*, **6**, 189-203.
- LAENG, B., CHABRIS, C. F., & KOSSLYN, S. M. (2002). Asymmetries in encoding spatial relations. In K. Hugdahl & R. Davidson (Eds.), *Brain asymmetry* (2nd ed., pp. 303-339). Cambridge, MA: MIT Press.
- LAENG, B., & TEODORESCU, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, **26**, 207-231.
- LEY, R. G. (1979). Cerebral asymmetries, emotional experience, and imagery: Implications for psychotherapy. In A. A. Sheikh & J. T. Shaffer (Eds.), *The potential of fantasy and imagination* (pp. 41-65). New York: Brandon House.
- LEVINE, D. N. (1982). Visual agnosia in monkey and man. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 629-670). Cambridge, MA: MIT Press.
- MARINKOVIC, K. (2004). Spatiotemporal dynamics of word processing in the human cortex. *Neuroscientist*, **10**, 142-152.
- MCDERMOTT, K. B., PETERSEN, S. E., WATSON, J. M., & OJEMANN, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia*, **41**, 293-303.
- MELLET, E., BRICOGNE, S., CRIVELLO, F., MAZOYER, B., DENIS, M., & TZOURIO-MAZOYER, N. (2002). Neural basis of mental scanning of a topographic representation built from a text. *Cerebral Cortex*, **12**, 1322-1330.
- MELLET, E., TZOURIO, N., CRIVELLO, F., JOLIOT, M., DENIS, M., & MAZOYER, B. (1996). Functional anatomy of spatial mental imagery generated from verbal instructions. *Journal of Neuroscience*, **16**, 6504-6512.
- MELLET, E., TZOURIO, N., DENIS, M., & MAZOYER, B. (1995). A positron emission tomography study of visual and mental spatial exploration. *Journal of Cognitive Neuroscience*, **7**, 433-445.
- MESULAM, M.-M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology*, **10**, 309-325.
- O'REGAN, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, **46**, 461-488.

- O'REGAN, J. K., & NOË, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral & Brain Sciences*, **24**, 939-1031.
- PLONER, C. J., GAYMARD, B. M., RIVAUD-PECHOUX, S., BAULAC, M., CLEMENCEAU, S., SAMSON, S., & PIERROT-DESEILLIGNY, C. (2000). Lesions affecting the parahippocampal cortex yield spatial memory deficits in humans. *Cerebral Cortex*, **10**, 1211-1216.
- PODGORNY, P., & SHEPARD, R. N. (1978). Functional representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 21-35.
- POSNER, M. I., & PETERSEN, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, **13**, 25-42.
- ROTA KOPS, E., HERZOG, H., SCHMID, A., HOLTE, S., & FEINENDEGEN, L. E. (1990). Performance characteristics of an eight-ring whole body PET scanner. *Journal of Computer Assisted Tomography*, **14**, 437-445.
- SERGEN, J. (1989). Image generation and processing of generated images in the cerebral hemispheres. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 170-178.
- SERGEN, J. (1990). The neuropsychology of visual image generation: Data, method, and theory. *Brain & Cognition*, **13**, 98-129.
- SERGEN, J., OHTA, S., & MACDONALD, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain*, **115**, 15-36.
- SERGEN, J., ZUCK, E., LEVESQUE, M., & MACDONALD, B. (1992). Positron emission tomography study of letter and object processing: Empirical findings and methodological considerations. *Cerebral Cortex*, **2**, 68-80.
- SLOTNICK, S. D., MOO, L. R., TESORO, M. A., & HART, J. (2001). Hemispheric asymmetry in categorical versus coordinate visuospatial processing revealed by temporary cortical deactivation. *Journal of Cognitive Neuroscience*, **13**, 1088-1096.
- SMITH, E. E., & JONIDES, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences*, **95**, 12061-12068.
- SMITH, E. E., & JONIDES, J. (1999). Storage and executive processes in the frontal lobes. *Science*, **283**, 1657-1661.
- TALAIRACH, J., & TOURNOUX, P. (1988). *Co-planar stereotaxic atlas of the human brain* (M. Rayport, Trans.). New York: Thieme.
- TANAKA, K., SAITO, H., FUKADA, Y., & MORIYA, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, **66**, 170-189.
- THOMAS, N. J. T. (1999). Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cognitive Science*, **23**, 207-245.
- THOMAS, N. J. T. (2002). The false dichotomy of imagery. *Behavioral & Brain Sciences*, **25**, 211.
- THOMPSON, W. L., KOSSLYN, S. M., SUKEL, K. E., & ALPERT, N. M. (2001). Mental imagery of high- and low-resolution gratings activates area 17. *NeuroImage*, **14**, 454-464.
- TIPPETT, L. (1992). The generation of visual images: A review of neuropsychological research and theory. *Psychological Bulletin*, **112**, 415-432.
- TULVING, E., KAPUR, S., CRAIK, F. I. M., MOSCOVITCH, M., & HOULE, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Sciences*, **91**, 2016-2020.
- TURKELTAUB, P. E., EDEN, G. F., JONES, K. M., & ZEFFIRO, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, **16**, 765-780.
- UNGERLEIDER, L. G., & MISHKIN, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press.
- WORSLEY, K. J., EVANS, A. C., MARRETT, S., & NEELIN, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, **12**, 900-918.
- WRAGA, M., THOMPSON, W. L., ALPERT, N. M., & KOSSLYN, S. M. (2003). Implicit transfer of motor strategies in mental rotation. *Brain & Cognition*, **52**, 135-143.

(Manuscript received December 19, 2003;
revision accepted for publication October 7, 2004.)