

Visual mental imagery and visual perception: Structural equivalence revealed by scanning processes

GREGOIRE BORST AND STEPHEN M. KOSSLYN
Harvard University, Cambridge, Massachusetts

The research reported in the present article investigates whether information is represented the same way in both visual mental imagery and the early phases of visual perception. In Experiment 1, the same participants scanned over patterns of dots in a mental image (with images based on a just-seen pattern), during perception, and in an iconic image. The time to scan increasing distances increased at comparable rates in the three tasks. However, in Experiment 2, when mental images were created from information stored in long-term memory, participants scanned more slowly in the mental image condition. Nevertheless, the rates of scanning in the perceptual tasks were highly correlated with the rates of scanning in the imagery tasks in both experiments. The results provide evidence that mental images and perceived stimuli are represented similarly and can be processed in the same way.

Mental images arise from perceptual representations that are created from stored information—not from information currently being registered by the senses. According to one theory, such images arise “when a representation of the type created during the initial phases of perception is present but the stimulus is not actually perceived” (Kosslyn, Thompson, & Ganis, 2006, p. 4). And, in fact, many previous studies have documented functional similarities between visual mental imagery and perception; these studies have relied on comparing behavior during imagery and perception tasks (e.g., Denis, 1991; Finke, 1985; Kosslyn, 1980; Paivio, 1986; Perky, 1910; Segal & Fusella, 1970; Shepard & Cooper, 1982), evaluating the effects of brain damage on how well patients perform the two types of tasks (e.g., Basso, Bisiach, & Luzzatti, 1980; Bisiach & Luzzatti, 1978; Farah, 1984; Farah, Levine, & Calvanio, 1988), and comparing recordings of activation in the brain while participants engage in the two types of tasks (e.g., Ganis, Thompson, & Kosslyn, 2004; Ghaëm et al., 1997; Ishai, Ungerleider, Martin, & Haxby, 2000; Kosslyn & Thompson, 2003; Kosslyn, Thompson, & Alpert, 1997; Mellet et al., 2000).

However, some findings have challenged the claim that visual mental imagery and visual perception rely on common underlying representations. For example, Behrmann, Winocur, and Moscovitch (1992) reported that a brain-damaged patient with a left homonymous hemianopia and a possible bilateral thinning of the occipital lobes (as revealed by positron emission tomography and magnetic resonance imaging) had disrupted object recognition but intact visual mental imagery. Moreover, Guariglia, Padovani, Pantano, and Pizzamiglio (1993) reported the

reverse dissociation in a patient who had a large lesion involving the right frontal lobe and the anterior temporal lobe. Denis and Kosslyn (1999), Ganis et al. (2004), Kosslyn (1994), and others (see Craver-Lemley & Reeves, 1987, 1992) have suggested that this double dissociation between visual imagery and visual perception arises because forming an image relies on top-down processes that are not always necessary in perception, whereas visual perception relies on bottom-up organizational processes that are not required in visual imagery. Nevertheless, we note that if a familiar object is seen under degraded conditions (i.e., an object is partially occluded or an object is seen under poor lighting conditions; see Ganis, Schendan, & Kosslyn, 2007), then top-down processing is likely to be used in visual perception. In such circumstances, visual mental imagery would not necessarily be more disrupted than perception following damage to brain areas involved in top-down processing.

These conjectures are consistent with the fact that not all of the same brain areas are activated during visual mental imagery and visual perception (Ganis et al., 2004; Kosslyn et al., 1997). For example, Ganis et al. (2004) found less overlap in activation during imagery and perception in the occipital and temporal lobes than in the frontal and parietal lobes—a result that they interpreted as indicating that perception relies in part on bottom-up organizational processes that are not used as extensively in imagery.

In the present study, we focused not on *functional* overlap between imagery and perception (which centers on how the content of the stored information is processed regardless of the format of the representation itself), but on possible *structural* overlap between the underlying rep-

representations (which focuses on the nature of the representations). Specifically, we investigated the possibility that visual imagery and perception rely on the same type of representation of the spatial layout of surfaces. Hypothesizing that visual imagery and perception rely on the same type of spatial representation is plausible in light of facts about the neural architecture of the visual system. In particular, the first cortical areas that process visual input are topographically organized: The spatial layout of the surfaces of objects is represented by the spatial layout of the patterns of activation on the cortex (see, e.g., Tootell, Silverman, Switkes, & De Valois, 1982). Thus, we can ask whether there is evidence that the same spatial properties constrain the processing of visual mental images and visual percepts.

In this study, we relied on an image-scanning paradigm that was introduced by Finke and Pinker (1982) and refined by Borst, Kosslyn, and Denis (2006). In this task, participants see a pattern of dots. After a short period of time, the dots disappear, at which point an arrow appears and the participant indicates whether the arrow would have pointed at a dot (if the dots were present as they had appeared initially). Researchers typically find that the time to make this decision increases with the distance between the arrow and dot—a result that has been taken to reflect image scanning (as is also subjectively reported). A crucial characteristic of this paradigm is that participants are never instructed to form or scan a mental image at any time; thus, the scanning effect (i.e., a linear increase in response times [RTs] with increasing distances) cannot be attributed to either experimenter expectancy effects or task demand effects (see Kosslyn et al., 2006).

The imagery scanning paradigm was designed to reveal information about the spatial structure of the representations that are processed. That is, in image scanning paradigms, RTs are used as a kind of “mental tape measure” to assess structural properties of the underlying representation. If mental images are in some sense pictorial, then space in the representation should embody actual space. If so, then the time to scan from one point to another on an object in a mental image should reflect the distance between the points. Consequently, as the distances between the points increase, RTs should also increase. Numerous researchers (e.g., Borst et al., 2006; Denis & Cocude, 1989; Dror & Kosslyn, 1994; Finke & Pinker, 1982, 1983; Kosslyn, 1973; Kosslyn, Ball, & Reiser, 1978; Pinker, 1980; Pinker, Choate, & Finke, 1984) have reported a linear increase in RTs with increasing distances scanned over objects in mental images. The conclusion of these studies was that the spatial structure of the representations underlying imagery affects the way that they are processed. Thus, the behavioral data have been used to infer structural properties of the representation itself—not simply how it functions to improve memory, interferes with certain sorts of processing, or the like.

However, researchers who subscribe to a propositional theory (e.g., Pylyshyn, 1973) have proposed a different interpretation of the scanning effect. According to this view, the results in scanning experiments reflect the amount of material that must be traversed in a set of linked asser-

tions; this view predicts that metric distance per se would not affect scanning times—a prediction that was shown to be incorrect by Kosslyn et al. (1978) and others (e.g., see Kosslyn, 1994). A second counterinterpretation of the scanning results relies on the “tacit knowledge” that participants have about visual processes (Pylyshyn, 1981). In particular, people may believe that visual scanning between successive fixation points takes more time with longer distances between the points. If so, then such knowledge may guide their responses in image scanning tasks, and the RTs would say nothing about the nature of the underlying representation. The present investigation bears on this theory.

An underlying assumption in image-scanning studies is that the same, relatively low level type of representation is used in imagery and perception (cf. Finke, 1985). If the same perceptual representations are used in visual mental imagery and perception, then one would expect similar results when participants scan an object in an image and when they scan the object during perception. However, this comparison has never been studied systematically or in depth. If similar results are obtained in the two conditions, then such findings would add a new piece of evidence that imagery does, in fact, rely on representations that depict information; by “depict,” we mean representations in which: (1) each part of the representation corresponds to a part of the represented object, so that (2) the distances between representations of the parts (in a representational space) correspond to the distances between the parts on the object itself. Depictive representations are qualitatively distinct from the sorts of descriptive representations that underlie language (see, e.g., Pylyshyn, 1973, 2002, 2003; for a review, see Kosslyn et al., 2006).

Although Beech (1980), Denis and Cocude (1989), Intons-Peterson (1983), and Jolicœur and Kosslyn (1985) compared image scanning and visual scanning, none of these studies compared image scanning and scanning an iconic image. Given that iconic image representations indisputably arise at a low level of processing in the visual system, showing strong similarities between scanning mental images and scanning iconic images would provide important information about the nature of mental image representation. Moreover, such a finding would further demonstrate that mental imagery is a distinct type of representation, as opposed to what Pylyshyn (2002, 2003) claimed.

In addition, none of the previous studies examined imagery and perceptual scanning (both during free view and based on iconic images) in the same participants. This comparison is crucial if we are to compare the scanning times in the two tasks. In the present studies, we examined whether the two types of scanning times are correlated; if the mental image representations share the spatial properties of the perceptual representations of the corresponding objects, then we would expect a high correlation between the slopes (increases in time with distance) in the different tasks. If a person scans quickly in one task, then he or she should scan quickly in the others. Furthermore, if a person scans slowly in one task, then he or she should scan slowly in the others. The goal of Experiment 1 was to compare the scanning abilities of the same group of

participants during visual perception, iconic memory, and visual mental imagery.

EXPERIMENT 1

We administered three scanning tasks. First, in the mental image scanning task (hereafter referred to as the *MI task*), a pattern of dots was presented on each trial (this was done in order to keep the procedures in the three tasks as comparable as possible, even though the participants had memorized this pattern earlier); then, a mask was flashed to eliminate any iconic image of the dots. Finally, an arrow was presented. Participants visualized the pattern of dots to decide whether the arrow pointed at a location that had been previously occupied by one of the dots. The backward masking procedure we used has been shown in many previous studies to eliminate iconic images (see, e.g., Becker, Pashler, & Anstis, 2000; Coltheart, 1983; Enns & Di Lollo, 2000; Gegenfurtner & Sperling, 1993).

Second, in the free visual scanning task (hereafter referred to as the *FV task*), a pattern of dots and an arrow remained on the screen until the participants decided whether the arrow pointed at one of the dots; thus, no mental image was required to perform this task. The FV task served in part as a “no mental image” control condition.

Third, in the iconic image scanning task (hereafter referred to as the *II task*), a pattern of dots with an arrow was presented very briefly, and participants scanned the distances based on an iconic image. The pioneering work of Sperling (1960) and subsequent studies (e.g., Avons & Phillips, 1980; Francis, 1996; Wede & Francis, 2006) have shown that the brain has the capacity to store for a short time (a few hundred milliseconds) a large amount of visual information. Given their short duration, residual iconic images should be able to support perceptual scanning briefly. We administered this task primarily to allow us to consider the possible “level” at which image scanning occurs in the nervous system. This task was also intended to control for possible effects of eye movements on the scanning rate (see Bahill & Stark, 1979; Fuchs, 1976). On each trial, participants were instructed to maintain their gaze on a fixation point in the center of the screen during the presentation of the pattern of dots. Posner, Nissen, and Ogden (1978) reported that when asked to maintain their gaze on a fixation point, participants were able to prevent eye movements on 94% of the trials. In addition, because saccades take between 150 msec and 350 msec to initiate (Kelley, Serences, Giesbrecht, & Yantis, 2008; Wilimzig, Schneider, & Schöner, 2006), presenting the stimuli for only 250 msec would have allowed—at most—a single eye movement.

In all three tasks, we recorded RTs and error rates (ERs). In order to compare the results obtained in the three scanning tasks, we matched the materials used in the three tasks (equivalent spatial relations between the dots, equivalent distances to scan, and equivalent arrows).

Method

Participants

Thirty-six volunteers from Harvard University and the local community participated in this study (18 females and 18 males). All

participants received either pay or course credit. Their average age was 22 years, 4 months; 31 were right-handed, 5 were left-handed. All reported normal or corrected-to-normal vision. Data from 4 additional people were not analyzed because 2 reported having followed the instructions less than 75% of the time, and 2 performed at least one of the tasks at chance levels of performance. All of the participants provided written consent and were tested in accordance with national and international norms governing the use of human research participants. The research was approved by the Harvard University Institutional Review Board.

Materials

We created three configurations of four black dots that were 7 mm in diameter (0.5° visual angle) in a 19×19 cm ($14.4^\circ \times 14.4^\circ$ visual angle) white square surrounded by a black frame. Pattern 2 was created by vertically flipping Pattern 1, and Pattern 3 was created by rotating Pattern 1 20° clockwise and then flipping it horizontally. A set of 48 arrows—2 cm in length (1.7° visual angle)—was also created; 24 arrows pointed directly at the center of one of the dots and 24 missed all of the dots by more than 40° . Each arrow was placed at one of six possible distances from the target (or nearest target) dot. They ranged from 1.5 to 9 cm, with 1.5-cm increments of differences in distance. For each dot, six arrows pointed at it for each of the six different distances. Because we were concerned that scanning along the horizontal or the vertical axis could affect the rate of scanning—especially in the perceptual conditions—no arrows were strictly horizontal or vertical. Each of the four arrows for a given distance was oriented at one of four possible angles (20° , 40° , 60° , or 80°). In some cases, the exact angles could not be achieved for a given arrow; hence, the angle could vary $\pm 5^\circ$ from the four values. For example, arrows at 20° could be oriented from 15° to 25° relative to the horizontal axis.

We chose the locations of the 24 *yes* arrows (which pointed at a dot) so that the six distances to be scanned were independent of the angles of disparity between the direction of the arrow (pointing at a target dot) and the nearest alternative dot. The correlation between the distances and the angles was nonsignificant [$r(22) = .22$]. Thus, as the distance increased between the tip of the arrow and the target dot, alternative dots did not become more crowded. Consequently, because we designed the arrows to preclude “perceptual crowding” (Pylshyn, 2002), such an explanation could not account for an increase in RTs with increasing distances. Similarly, for the *no* arrows (which missed all of the dots by more than 40°), when the distances between the arrows and the dots increased, the angle with which the arrows missed the dots did not systematically vary [$r(14) = -.17$]. This correlation was calculated only on 16 of the 24 *no* arrows, because 8 of them missed the dots by more than 90° ; hence, no distance could be associated with them.

Finally, we placed the arrows and dots within a virtual circle with a 9-cm radius in order to discourage the participants from using the black frame as a reference for memorizing the positions of the dots. The set of arrows was submitted to the same spatial transformation (rotation and flipping) as the one used to create the corresponding patterns of dots, which enabled us to have the exact same distances for each pattern of dots. The stimulus patterns are illustrated in Figure 1. Stimuli were presented on a 17-in. monitor with resolution of $1,280 \times 1,024$ pixels and a refresh rate of 75 Hz.

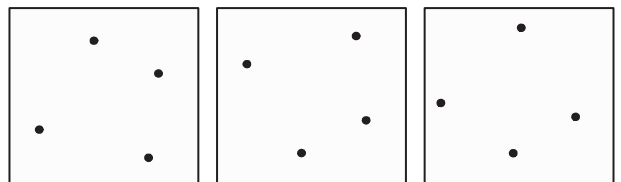


Figure 1. Patterns of dots used in the scanning tasks.

Procedure

The participants were tested individually, sitting 75 cm from a computer screen. Each participant performed the MI, FV, and II tasks. A given participant performed the three conditions with a single pattern of dots on all trials in a given task. However, the pattern of dots was different in each of the tasks. We counterbalanced over participants which configuration of dots was used for each task.

In each task, participants first performed a practice block of 48 trials. The computer provided feedback, and trials on which there were errors were repeated at the end of the block. Then, in each task, six blocks of 48 experimental trials were presented with no feedback, and trials on which there were errors were repeated at the end of each block. The order of the trials was randomized, except that no more than three *yes* or three *no* trials could occur in a row. The onset of the arrow started a timer, which was stopped when one of the two response keys was pressed. RTs and the nature of the response were recorded.

Because we were interested in the correlations among slopes, we needed to test the participants in the same experimental situation. However, to compare the mean slopes of the best-fitting lines, we had to counterbalance the order of the tasks. In a pilot study, we observed that the participants' scanning efficiency in the II task correlated most strongly with their scanning efficiency in the MI task; thus, to avoid possible transfer of general processes between these tasks (which potentially could have accounted for the correlations), we chose to administer the FV task between the two tasks that were exhibiting the strongest correlation in the pilot study. Thus, we assigned participants to one of two counterbalancing groups. In

Group 1 (18 participants), the order was MI, FV, and II; in Group 2 (18 participants), this order was reversed.

Participants were asked to follow the written instructions displayed on the screen. A 5-min break was provided between each task. At the end of the experiment, participants completed a debriefing questionnaire to ensure that they had no idea of the purpose of the experiment and that they followed the instructions at least 75% of the time in each of the tasks.

MI task. We began by showing the participants the pattern of dots on a hard-copy printout, and we asked them to draw the locations of the dots from memory on a blank sheet of paper, with the empty 19×19 cm black frame printed on it. We printed a hard copy of each of the original patterns on transparency sheets. In order to compare their drawings with the original pattern, participants superimposed the transparency on their drawings. Participants were asked to correct their drawings without the transparency and to then superimpose the transparency a second time. The draw-and-study procedure was repeated until all dots were drawn within 0.35 cm of their actual location. Depending on the participant, 1–7 drawings were required to reach this criterion. The draw-and-study procedure was an attempt to ensure that all participants started the task having the same knowledge of the positions of the dots; thus, the participants' ability to memorize the dots during the course of the task would not affect individual differences in RTs or ERs. The participants were told that they would be cued to visualize this pattern.

On each trial of the task itself (as illustrated in Figure 2), a fixation point first appeared in the middle of the screen for 1 sec. The pattern of four dots then appeared for 2 sec, which cued the par-

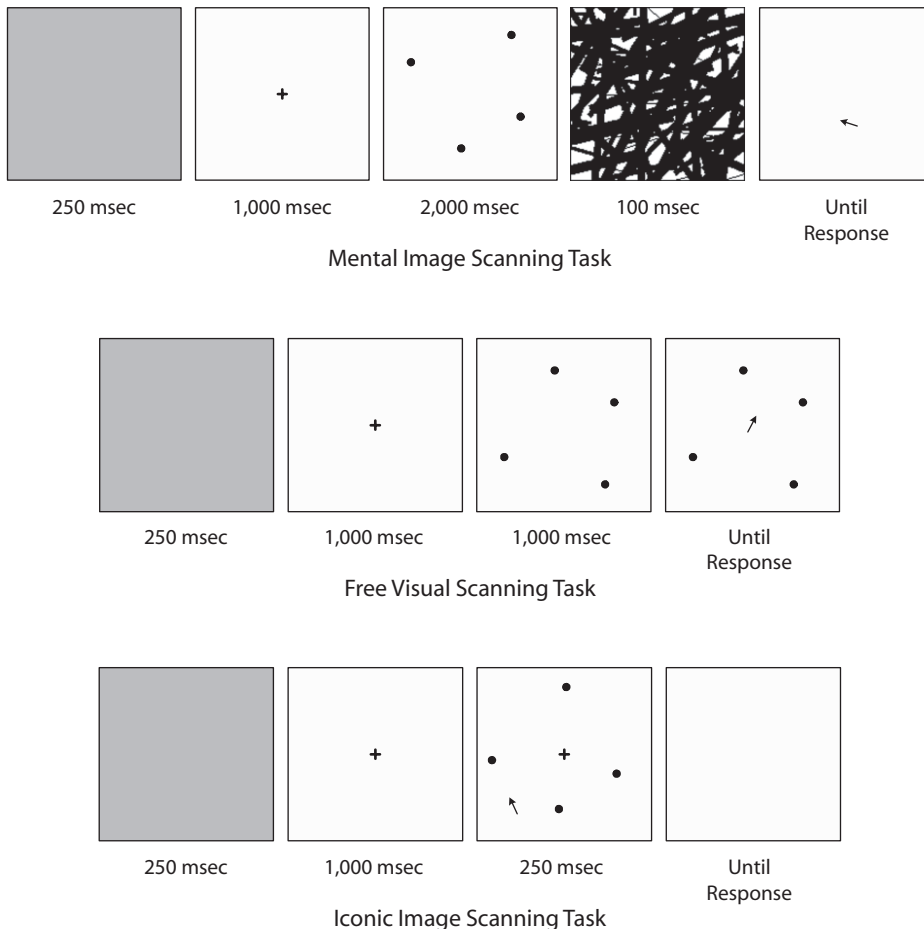


Figure 2. Experiment 1: The procedures used in the three scanning tasks.

ticipants to form the image after the mask (multiple black crossing lines on a white background; see Figure 2); the mask was displayed for 100 msec in order to disrupt any residual iconic image. Following this, an arrow appeared on the screen in the black frame and remained visible until the participants responded (see Figure 2). The participants were instructed to decide as quickly and accurately as possible whether the arrow pointed to a location previously occupied by one of the dots. If so, they were to press one key with their dominant hand (labeled with a “Y” on a green sticker); if not, they were to press the other key with their other hand (labeled with a “N” on a red sticker). At the beginning of the task, an example of a *yes* trial was presented by simultaneously displaying the pattern of dots and an arrow that pointed to one of them.

FV task. The task and procedure were the same as those in the MI task, except that the participants did not memorize the pattern in advance. The trial structure was as follows: On each trial, a fixation point was presented in the center of the screen for 1 sec; the pattern of four dots then appeared for 1 sec in the 19×19 cm black frame; an arrow appeared along with the dots in the frame and remained visible until one of the response keys was pressed (see Figure 2).

II task. The task and procedure were the same as those in the FV task, except that the pattern of dots appeared along with an arrow in the 19×19 cm frame, and this display was visible for 250 msec before it disappeared (but the outside square frame remained visible; see Figure 2). We presented the dots and the arrow simultaneously because we were concerned that, given the short duration of the presentation of the stimuli, displaying the arrow after the dots could have masked them. If so, this would have introduced spatial imprecision, which would in turn have contaminated our estimates of the scanning times.

Results

As a first step, we analyzed RTs and ERs to determine whether we replicated earlier findings of studies that used this scanning paradigm. Following this, we compared the steepness of the slopes of the best-fitting lines, the height of the intercepts, and the ERs in the three tasks. We then analyzed the relationship between individual differences in performance of the three tasks.

Preliminary analyses did not reveal any effect of—or interaction with—gender or the specific patterns. The order of the tasks also did not affect the results, except for the height of the intercept in the II task. This was higher in Group 2, which performed the II task first [$M = 509$ msec for Group 2, as compared with $M = 445$ msec for Group 1; $t(34) = 3.64, p < .005$]. We found no effect of the order of the tasks on the other measures. Although we did not use the full range of possible counterbalancing orders, the fact that we found virtually no differences between the most extreme orders—formed by switching the first and last tasks—is strong evidence that neither practice nor fatigue greatly affected the results. Thus, we pooled the data for the two counterbalancing groups, as well for males and females and for the different patterns, and we will not address these factors in the following description of the results.

Analysis of RTs and ERs

We analyzed separately the RTs from correct responses on *yes* and *no* trials, because we expected participants to scan the entire distances in the *yes* trials, but not always to do so in the *no* trials; that is, when it became obvious that the trajectory was going to miss a dot, participants may simply have stopped scanning, and we had no way to estimate at what point such a judgment may have been

made. Prior to the analyses, we eliminated outliers, which were defined as RTs greater than 2 *SDs* from the mean of that distance for that participant. In the *yes* trials, we also considered RTs under 250 msec as outliers, because such times could not have reflected the cognitive processes of interest (i.e., the scanning processes). Outliers occurred on 4.2% of the trials in the MI task, and on 3.1% in the two visual scanning tasks. Finally, trials on which participants made errors were repeated on average 1.11 times.

Yes trials. In order to discover whether we replicated previous findings, we first averaged the RTs over the trials for each distance in each task for each participant; then, we conducted separate ANOVAs on data from the three tasks. In the MI task, we found that different distances between the tip of an arrow and the previous location of a target dot required different amounts of time to scan [$F(5,175) = 13.47, p < .0005$]. As is evident in Figure 3, the best-fitting linear function calculated by the method of least squares revealed that RTs increased linearly with distance [$F(1,35) = 62.89, p < .0005$]. The Bravais–Pearson correlation between times and distance was $r(4) = .93, p < .01$. These results replicated those reported by Finke and Pinker (1982, 1983) and suggest that participants did use image scanning to perform this task.

We conducted the same analysis on the data from the FV task and found that RTs varied for the different distances [$F(5,175) = 56.55, p < .0005$]. Moreover, RTs increased linearly with increasing distance [$F(1,35) = 125.07, p < .0005$]; as is shown in Figure 3, RTs were correlated with distance [$r(4) = .99, p < .01$]. These results are consistent with those reported by previous studies, even though we used a different scanning paradigm, which confirmed the robustness of the scanning effect (Beech, 1980; Denis & Cocude, 1989; Pinker, 1980).

Finally, in the II task, we processed the data as follows: We included the 250 msec of presentation of the stimulus in the RTs to take into account the possibility that some participants could have started to scan before the offset of the stimulus. In addition, we did not include RTs from trials in which participants responded in less than 250 msec,

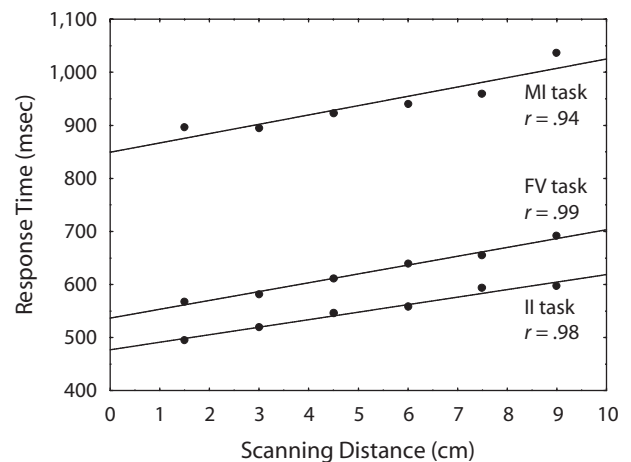


Figure 3. Experiment 1: The times to scan increasing distances in the three scanning tasks.

which allowed us to remove trials in which scanning could have been completed while the stimulus was present. RTs again varied with distance [$F(5,175) = 79.1, p < .0005$], and RTs increased linearly with increasing distance [$F(1,35) = 173.88, p < .0005$]. As is illustrated in Figure 3, we again found a strong positive correlation between scanning times and distance [$r(4) = .98, p < .01$]. These results are of great interest not only because they provide evidence that even with limited eye movements (at most, one saccade), participants were able to scan the distances between the tip of an arrow and a target dot, but also because they provide a first hint that mental image representations may be similar to iconic image representations. It should be noted that the participants reported during the postexperiment debriefing that they did not use mental images in the two perceptual tasks (FV and II).

No trials. We excluded from our analysis 8 of the 24 arrows because they missed the dots by more than 90° , and we could not associate the arrows with a specific distance. The 16 remaining arrows were assigned to one of two groups of 8 arrows: Half were less than 4 cm from the nearest *yes* dot and half were more than 5 cm from the nearest *yes* dot. There was no effect of distance on the *no* RTs in any of the three tasks [$F(1,35) < 1$, n.s., in the MI task; $F(1,35) < 1$, n.s., in the FV task; and $F(1,35) < 1$, n.s., in the II task]. The lack of effect of distance on the RTs in the *no* trials is probably a result of the fact that we designed the arrows to miss all the dots by more than 40° ; thus, the discrimination was very easy. Participants were faster on the *no* trials than on the *yes* trials, both in the II task (with means of 306 vs. 551 msec, respectively) [$t(35) = 19.92, p < .0001$] and in the MI task (862 vs. 942 msec, respectively) [$t(35) = 4.25, p < .0005$]. However, in the FV task, participants were not significantly faster on the *no* trials than they were on the *yes* trials (with means of 607 vs. 624 msec) [$t(35) = 1.09$, n.s.]. Thus, in the MI task and the II task, given the angles chosen, participants did not necessarily need to scan to decide whether the arrow missed the dots; we suggest that a fast attentional process (distinct from imagery) could be sufficient to make the necessary discrimination. Given that RTs did not increase with distance in the *no* trials for the FV trials, we cannot infer that participants scanned during these trials; instead, the participants may simply have “double-checked” the *no* trials when the pattern was in free view in order to ensure that the arrow did, in fact, miss the dots. If so, they would not be faster for *no* trials, but they also would not require more time with longer distances. However, if a smaller angle had been chosen, we would have probably observed an increase of the RTs with increasing distance (as reported by Finke & Pinker, 1982, 1983).

Slopes. To examine the similarity of the scanning processes in the three tasks, we directly compared the slopes of the best-fitting lines from each of them. Thus, we calculated the slope for each participant in each task and submitted these slopes to an ANOVA. The mean slopes in all three tasks were comparable [$F(2,70) = 1.74$, n.s.; the mean for MI = 16 msec/cm, FV = 17 msec/cm, and II = 14 msec/cm]. In short, we had evidence that participants scanned at the same rate during imagery and perception. We found these results surprising, because iconic images

are processed at a low level in the visual system; thus, the processes that operate on this representation should be more “automatic” and hence faster than the “controlled” processes that operate on visual mental images.

Intercepts. In addition, as is evident in Figure 3, an analysis of the height of the intercepts revealed differences among the tasks [$F(2,70) = 121.10, p < .0005$]. The specific comparisons indicated that the mean intercept was higher in the MI task ($M = 864$ msec) than in the FV task ($M = 537$ msec) [$t(35) = 10.96, p < .0005$] or in the II task ($M = 477$ msec) [$t(35) = 11.67, p < .0005$]. Finally, the mean intercept also differed between the two perceptual tasks [$t(35) = 4.75, p < .0005$]. As was noted when we discussed results from the II task, in order to account for the fact that participants could have started to scan before the offset of the stimulus, the presentation time of the display (250 msec) was included in the RTs. Consequently, the average height of the intercepts in the II task was, in fact, 227 msec ($M = 477 - 250$ msec). This result is consistent with the duration of iconic memory as reported in the original research of Sperling (1960) and subsequent studies (e.g., Avons & Phillips, 1980; Francis, 1996; Wede & Francis, 2006).

The difference between the heights of the intercepts in the II task and the MI task is crucial because it demonstrates that the backward masking procedure did, in fact, eliminate iconic images in the MI task; only if iconic images are eliminated would participants need to create a mental image in order to scan. If iconic images were scanned in both tasks, then the heights of the intercepts should have been the same—but they were not. Moreover, iconic images are retained for only a few hundred milliseconds, but participants started to scan on average 864 msec after the mask in the MI task; this finding also speaks against the participants performing the MI task by scanning an iconic image.

ERs. Finally, we compared the ERs on the *yes* and the *no* trials in the three tasks. Participants made different numbers of errors in the different tasks [$F(2,70) = 25.10, p < .0005$]. Specifically, they made more errors in the MI task ($M = 6.4\%$) than in the FV task ($M = 3\%$) [$t(35) = 5.87, p < .0005$]. In the FV task, the pattern of dots and the arrow were presented simultaneously; thus, the judgments were easier than they were when the pattern of dots was not physically present on the screen. Furthermore, participants made more errors in the II task than in the FV task [$t(35) = 6.46, p < .0005$]. One explanation for this finding is that the II task required greater attention than the FV task because of the brief time that the dots and the arrow were presented (250 msec). However, the participants made comparable numbers of errors in the MI and II tasks ($M = 6.2\%$) [$t(35) = 1.39$, n.s.].

Correlational Analyses

In order to consider whether the same scanning processes were tapped in each of the three tasks—as was suggested by the analysis of the average RTs and ERs—we examined the correlation between all dependent variables. If the same underlying scanning processes—applied to the same types of representations—were used in all three

tasks, then we would expect high correlations between the slopes. For each participant, we considered four measures for each of the three scanning tasks: (1) the slope of the best-fitting lines, (2) the intercept, (3) the correlation coefficient between RTs and distance, and (4) the ER. We checked split-half reliability for each measure. Critically, the slope measures were all reliable [MI, $r(34) = .73$; FV, $r(34) = .82$; II, $r(34) = .79$, all $ps < .01$]. Reliability coefficients for the intercept and the ERs ranged from .90 to .99 (all $ps < .01$). However, we found low reliability for the correlation coefficients between RT and distance [$r(34)$ ranging from .26 to .38]. The lack of reliability of the coefficient of correlations between RTs and distances may have simply reflected the higher sensitivity of this index to outliers. Given this finding, we did not analyze these correlations further. The correlations between all dependent variables are presented in Table 1.

Slopes. We next obtained the correlation coefficients for the slopes of the increases in RT with distance in the three tasks: MI and FV [$r(34) = .59, p < .01$], MI and II [$r(34) = .62, p < .01$], and FV and II [$r(34) = .70, p < .01$]. We then calculated the correlation coefficients corrected by attenuation [$r_{xy'} = r_{xy} / \sqrt{(r_{xx}r_{yy})}$; see Spearman, 1907], which is an estimate of what could be the relationship between two tasks if the two measures were perfectly reliable, and we found the following correlations: MI and FV [$r(34) = .76, p < .01$], MI and II [$r(34) = .81, p < .01$], and FV and II [$r(34) = .87, p < .01$]. Finally, we used Williams's (1959) modified Hotelling t test to compare the correlation coefficients between the slopes in the three scanning tasks. The correlation (corrected by attenuation) between the slopes in the two perceptual tasks was no stronger than the correlation between the MI and II slopes [$t(33) = 1.11, n.s.$]. However, the correlation between the FV and II slopes was stronger than the correlation between the MI and FV slopes [$t(33) = 2.09, p < .025$].

The pattern of correlations suggests that although the participants' scanning speed in the mental imagery task is predicted well by their scanning speed over an iconic image, the process of scanning a pattern in free view is affected by at least some different factors than those used to scan mental images and iconic images. This

finding is consistent with our findings from *no* trials, which were noted earlier.

Intercepts. The intercepts were also related, with the following correlations: MI and FV [$r(34) = .57, p < .01$], FV and II [$r(34) = .43, p < .05$], and II and MI [$r(34) = .35, p < .05$]. When corrected for attenuation, we found the following correlations: MI and FV [$r(34) = .61, p < .01$], FV and II [$r(34) = .48, p < .01$], and II and MI [$r(34) = .37, p < .05$]. However, the intercepts and the slopes were not related in the FV task [$r(34) = .27, n.s.$] or in the II task [$r(34) = .24, n.s.$]. For the MI task, the correlation between slopes and intercepts just reached significance [$r(34) = .43, p < .05$].

Finally, using Steiger's (1980) equation—based on Fisher's (1921) z transformation of the coefficient of correlation—the correlation (corrected by attenuation) between the slopes in the MI and II tasks was stronger than between the correlation (corrected by attenuation) between the intercepts in these two tasks ($z = 2.97, p < .005$). However, the same comparison of the correlation between the slopes and the intercepts on the MI task and the FV task revealed no significant difference ($z = 1.54, n.s.$).

On the basis of this pattern of correlations, we infer that the intercepts reflect the participants' efficiency on common general processes that cannot account for participants' scanning efficiency (which is reflected by the slopes).

ERs. We also correlated the error rates in the three scanning tasks: MI and FV [$r(34) = .51, p < .01$], MI and II [$r(34) = .64, p < .01$], and FV and II [$r(34) = .73, p < .01$]. When corrected by attenuation, the correlations were: MI and FV [$r(34) = .53, p < .01$], MI and II [$r(34) = .67, p < .01$], and FV and II [$r(34) = .74, p < .01$]. In addition, we observed correlations between the ERs and the height of the intercepts in the two perceptual tasks [$r(34) = -.47, p < .01$, for the FV task, and $r(34) = -.49, p < .01$, for the II task]. Thus, these two indices were affected by a speed–accuracy trade-off. However, in the MI task, ERs and the height of the intercepts were not correlated [$r(34) = .13, n.s.$], which ruled out a speed–accuracy trade-off on the intercepts.

Finally, we did not find significant correlations between the ERs and the slopes of the best-fitting lines in any of the

Table 1
Experiment 1: Matrix of Correlations

	Task	Slope			Intercept			Coefficient			Error Rate		
		MI	FV	II	MI	FV	II	MI	FV	II	MI	FV	II
Slope	MI	1.00	.59**	.62**	.43**	.31	-.03	.27	.28	.33	-.17	-.32	-.24
	FV		1.00	.70**	.65**	.27	.07	-.09	.52**	.47**	-.07	-.31	-.25
	II			1.00	.52**	.50**	.24	.19	.25	.44**	-.07	-.38*	-.30
Intercept	MI				1.00	.57**	.35*	-.23	.21	.15	.13	-.40*	-.23
	FV					1.00	.43**	.07	-.19	-.08	.04	-.47**	-.32
	II						1.00	-.25	-.15	-.25	-.17	-.31	-.49**
Coefficient	MI							1.00	.18	.11	.04	.08	.11
	FV								1.00	.26	-.18	-.08	-.02
	II									1.00	.01	-.04	.09
Error rate	MI										1.00	.51**	.64**
	FV											1.00	.73**
	II												1.00

Note—Slope, slope of the best-fitting line; Intercept, intercept at the ordinate; Coefficient, coefficient of correlation between times and distances to scan; MI, the mental image scanning task; FV, the free visual scanning task; II, the iconic image scanning task. * $p < .05$. ** $p < .01$.

three tasks. Thus, the efficiency of the scanning process per se was not affected by a speed–accuracy trade-off.

Discussion

In all three tasks, the time to scan increased linearly when distance between the arrow and the dots increased on the trials in which arrows were actually pointing at target dots. We interpreted these results as reflecting the time to shift the point of attention across a spatial representation. These findings are of theoretical importance for at least three reasons: First, a theory that posits that scanning iconic images takes place at “lower” levels in the nervous system than scanning mental images would predict differences in slopes (in fact, we find the comparable slopes in the two conditions very counterintuitive). Second, a propositional theory—of the sort proposed by Pylyshyn (1973)—would not predict an increase in times with increasing distance in the mental imagery task, given that only metric distance is varied (and that “perceptual crowding,” which will be discussed shortly, cannot explain the results). Third, the “tacit knowledge” theory—later championed by Pylyshyn (1981)—would not predict comparable times to scan a mental image and an iconic image; the iconic image scanning condition in our experiment is a laboratory task that was never previously encountered by the participants; hence, they presumably would have no tacit knowledge of how to behave in this task. Moreover, postexperiment debriefing revealed that none of the participants expected a relationship between time to scan and distance in the iconic image scanning condition, nor were they aware of our interest in this relationship.

We wish to infer that the present results implicate a common representation that is processed in the three tasks. However, Pylyshyn (2002) proposed that “perceptual crowding” could account for the strong linear relationship between RT and distance in this imagery task. The argument is that as the distance increases between the arrow and target dot, alternative target dots become more crowded. If this is the case, then RTs increased not as an effect of scanning greater distances but because of the increased difficulty of discriminating the target dot from the remaining dots. An alternative version of this view is that the angle of the *no* trials and the nearest *yes* dots became smaller as distance increased, which would also make the discrimination (now between a *yes* vs. a *no* decision) increasingly difficult with increasing distance. Neither variant of this account is viable for our results, however, because we designed the stimuli so that the angle between the arrow direction and the nearest alternative dot did not become smaller with greater distance.

One could also argue that during the perceptual scanning tasks (free visual and iconic image scanning), participants were able to memorize the locations of the dots and, consequently, they performed the task without any visual stimulation. However, Finke and Pinker (1983) showed that when participants were told the location of the arrow in advance, they could perform accurately, but the scanning effect did not occur. The fact that we did find such scanning effects in the perceptual tasks indicates that the participants did process the spatial arrays by scanning

from the tip of the arrows to the location of the dots, presumably because the arrows appeared at unexpected location (following the procedure of Finke & Pinker, 1983, Experiment 1).

In Experiment 1, by showing not only that the slopes (reflecting the increase in the time to scan increasing distances) are comparable within the same participants, but also that scanning abilities are strongly correlated between the perceptual scanning conditions (FV and II tasks) and image scanning condition, we have provided evidence that the underlying representations share the same spatial structure.

However, one could also argue that the correlation between the slopes in the MI and FV task is an artifact of how participants moved their eyes. In fact, Brandt and Stark (1997) and Laeng and Teodorescu (2002) showed that eye movements occur during visual mental imagery and that the sequence of fixations during imagery is similar to the sequence of fixations during perception. However, when they compared the performance of participants in two different learning conditions (free eye exploration vs. fixation), Laeng and Teodorescu found no difference in participants’ performance. Thus, eye movements per se could not have been responsible for spatial memory (as noted by Mast & Kosslyn, 2002). Moreover, the fact that in our Experiment 1 the slopes were correlated to the same extent in both the MI versus FV comparison and the MI versus II comparison (where—at most—a single eye movement was possible), speaks against a major role of eye movements in the present tasks. Although saccade duration does increase with saccade distance (see, e.g., Rayner, 1998), given the range of distances scanned in our tasks, such an increase would not produce the longer RTs for longer distances we found in the II task.

In addition, if overall speed or general processes (reflected by the intercepts) were responsible for the correlations of the slopes, we would have expected the intercepts to be as highly correlated as the slopes. However, the correlation between the MI and the II slopes was higher than the one for the intercepts in these two tasks. Moreover, no significant correlations were found between the slopes and the intercepts in the perceptual scanning tasks. The correlation found between the slopes and the intercepts in the imagery task was probably due to the fact that the intercept in this task reflects the time to generate the image. In conclusion, we want to argue that the intercepts reflect general processes (such as the speed of information processing during encoding the patterns and producing responses), whereas the slopes of the best-fitting lines reflect the efficiency of scanning per se.

Of particular note, the correlations between the slopes provide evidence that the same process is used to scan iconic images and visual mental images—which implies that a comparable type of representation is processed in the two cases. That is, a given type of process is only appropriate for representations that have compatible characteristics. Iconic images are clearly represented spatially, hence, if a process appropriate for scanning such representations is also used in mental imagery, this implies that mental image representations share certain structural

features with iconic image representations. Given the fact that visual information persists in primary visual cortex (V1) after display offset (see, e.g., Duysens, Orban, Cremieux, & Maes, 1985; Engel, 1970; Supèr, Spekreijse, & Lamme, 2001), the relationship between the scanning processes in these two tasks—although surprising and counterintuitive—supports the hypothesis that visual mental images rely on relatively low-level representations in the visual system.

However, several aspects of the tasks could have biased the results, and would have led us to overestimate the structural equivalence between perceptual representation and mental image representation. Experiment 2 addresses the key issues.

EXPERIMENT 2

One could argue that the iconic image scanning task did not, in fact, require participants to scan an iconic image. First, because we presented the dots and the arrow simultaneously and presented these stimuli for 250 msec, some of the participants conceivably could have scanned the pattern of dots itself and not an iconic image. Second, because we used a single pattern of dots on all trials in the II task, participants could have relied more on their long-term memory of the position of the dots than on their iconic image, per se.

In addition, one could question the design of the mental image scanning task of Experiment 1. First, it is possible that the mask did not entirely eliminate the iconic image of the stimulus. Second, participants could make eye movements over the to-be-scanned pattern in the mental image scanning task, which could have affected the slopes of the increases in RT with distance.

In order to address these issues, in Experiment 2, we revised both the iconic image scanning task and the mental image scanning task. In the II task, we presented the pattern for 125 msec and immediately replaced it by an arrow that was shown for 125 msec. In order to control for the possibility that the participant could have memorized the pattern of dots in the II task, we presented a different pattern of dots on each of the trials in a given experimental block. In the MI task, in order to ensure that participants did not rely on their iconic memory (which would have been possible if the mask were inadequate), we no longer displayed the pattern of dots on each trial. Instead, the participants memorized the pattern of dots prior to the task and were asked to visualize the dots to decide whether an arrow would have pointed at one of them if they were on the screen as they appeared when memorized. Participants were instructed to keep their gaze focused on a fixation cross during the course of trials in both tasks in order to limit possible eye movements.

Method

Participants

Thirty-six volunteers from Harvard University and the local community participated in this study (24 females and 12 males). None of these volunteers participated in Experiment 1. All participants received either pay or course credit. Their average age was 23 years,

5 months; 35 were right-handed and 1 was left-handed. All reported normal or corrected-to-normal vision. Data from 5 additional people were not analyzed, because 2 reported having followed the instructions less than 75% of the time, and 3 performed at least one of the tasks at chance levels of performance. All of the participants provided written consent and were tested in accordance with national and international norms governing the use of human research participants. The research was approved by the Harvard University Institutional Review Board.

Materials

In the MI task, we used one of the three patterns of four black dots designed in Experiment 1 (Pattern 1, see Figure 1). In addition to the arrows used in Experiment 1, we designed five new sets of 48 arrows (24 *yes* arrows and 24 *no* arrows). The new arrows respected the same set of constraints as those in Experiment 1 (i.e., they had the same size, the same set of distances for the *yes* arrows, and the same properties to preclude a possible “crowding effect”; no arrows were strictly horizontal or vertical, and all arrows were placed within a virtual circle with a 9-cm radius).

In the II task, we created 48 different patterns of four dots with the same characteristics as those in Experiment 1. None of these patterns were similar to the pattern of dots used in the MI task. For each pattern, we designed six arrows: Three pointed directly at the center of one of the dots (*yes* arrows) and three missed all the dots by more than 40° (*no* arrows). Each of the *yes* arrows was placed at one of six possible distances from a target dot, ranging from 1.5 cm to 9 cm, with 1.5-cm increments in increasing distance. Respecting the same constraints used in Experiment 1, we designed the 144 *yes* trials to preclude a “perceptual crowding” effect (Pylyshyn, 2002) [$r(142) = .11$, n.s.]. The *no* arrows also were designed with the same constraints as those in Experiment 1; thus, when the distance between the arrows and the dots increased, the angle with which the arrows missed the dots did not systematically vary [$r(94) = .09$, n.s.]. (We restricted the analysis to 96 of the 144 *no* arrows for which we could assign a distance.) We presented the stimuli on the same computer screen used in Experiment 1, with the identical brightness and contrast settings.

Procedure

The procedure was identical to that used in the corresponding task in Experiment 1, except that participants performed practice blocks of 12 trials. We counterbalanced the order of presentation of the two tasks.

MI task. As in Experiment 1, we used a draw-and-study procedure that was repeated until all dots were drawn within 0.35 cm of their actual location in two consecutive trials. Participants required from 1 to 11 drawings to reach this criterion. Following learning, each test trial began with a fixation cross, which appeared in the middle of the screen for 2.5 sec. Participants were asked to visualize the dots at their exact locations, in the same form that they studied, without moving their eyes from the fixation cross. Following this, an arrow appeared on the screen in the black frame, remaining visible until the participants responded (see Figure 4). The participants were instructed to decide as quickly and as accurately as possible whether the arrow pointed to a location occupied by one of the dots they had memorized during the draw-and-study procedure, while keeping their gaze focused on the fixation cross (note that no reference to scanning was ever made). All other aspects of the procedure were the same as those of the MI condition of Experiment 1.

II task. As opposed to the MI task, the participants did not memorize the pattern in advance. On each trial, a fixation point was presented in the center of the screen for 1 sec; one of the 48 patterns of four dots then appeared for 125 msec in the 19 × 19 cm black frame; the pattern of dots then was replaced by an arrow in the black frame for 125 msec. Participants were instructed to keep their gazes focused on the fixation point in the middle of the screen at all times (see Figure 4). After the offset of the arrow, participants decided whether the arrow pointed at one of the dots. The participants used the same two keys that were used in the MI task to respond.

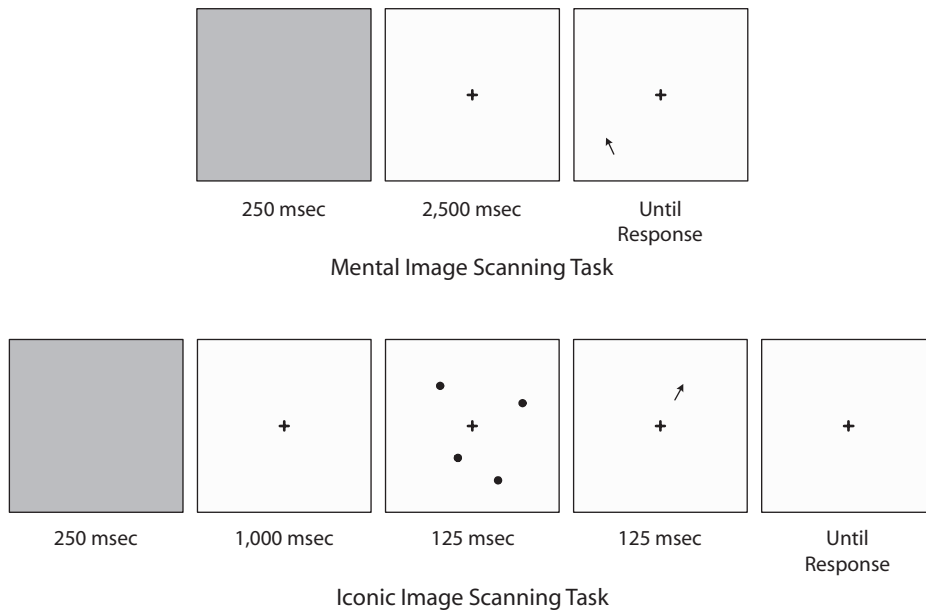


Figure 4. Experiment 2: The procedures used in the two scanning tasks.

On each block of 48 trials, we designed the arrows for each of the 48 patterns of dots so that for the 24 *yes* arrows, each of the six possible distances between the tip of the arrow and the dots appeared four times. However, a pattern of dots was presented only once per block and was never associated with the same arrow in the six experimental blocks of trials. The order of the trials was randomized, except that no more than three *yes* or three *no* trials could occur in a row. All other aspects of the procedure were the same as those in the II task of Experiment 1.

Results

As a first step, we analyzed RTs and ERs in the same way as we did in Experiment 1. Preliminary analyses did not reveal an effect of gender or of the order of the tasks on all three dependent variables in both scanning tasks. Thus, we pooled the data over these variables, and we will not address these factors in the following report of the results.

Analysis of RTs and ERs

We conducted identical analyses to those in Experiment 1. Defining outliers as in Experiment 1, there were 4.9% in the MI task and 2.8% in the II task. Trials on which participants made errors were repeated, on average, 1.25 times.

Yes trials. In the MI task, we found different times to scan different distances from the tip of an arrow to a location of a dot previously memorized [$F(5,175) = 13.07$, $p < .0005$]. In addition, as shown in Figure 5, RTs increased linearly with distance, as documented by the best-fitting function calculated by the method of least squares [$F(1,35) = 67.84$, $p < .0005$]. RTs and distance were highly correlated [$r(4) = .88$, $p < .01$]. These results replicated those obtained in Experiment 1 and the results reported by Finke and Pinker (1982, 1983). The results suggested that participants scanned their mental image of the pattern of dots to evaluate the trials.

We conducted the same analysis on the data from the II task and found that RTs varied for the different distances [$F(5,175) = 68.32$, $p < .0005$]. Moreover, RTs increased linearly with increasing distance [$F(1,35) = 192.36$, $p < .0005$]; as shown in Figure 5, RTs were highly correlated with distance [$r(4) = .95$, $p < .01$]. These results are consistent with those reported in Experiment 1, and they suggest that participants scanned their iconic images to determine whether an arrow was pointing at a dot.

No trials. In both tasks, we analyzed the data from 96 arrows for which we could assign a distance, and we processed the data in the same way as in Experiment 1. There was no effect of distance on the *no* RTs [for both tasks, $F(1,35) < 1$, n.s.]. As in Experiment 1, participants were faster on *no* trials in both the II task (with means of 433 vs.

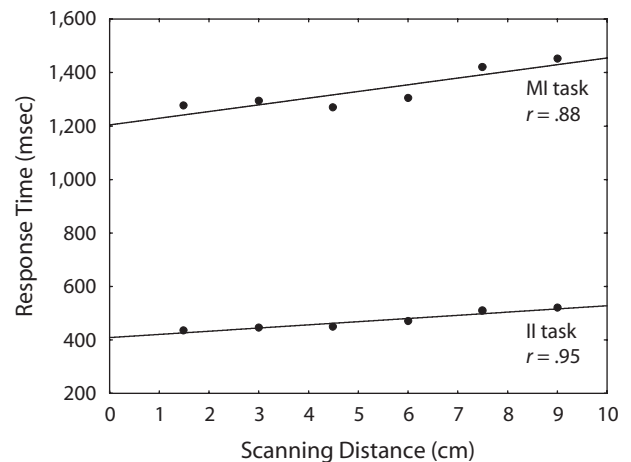


Figure 5. Experiment 2: The times to scan increasing distances in the two scanning tasks.

472 msec) [$t(35) = 2.52, p < .01$] and the MI task (with means of 971 vs. 1,336 msec) [$t(35) = 4.06, p < .0005$]; these results suggest that the participants did not need to scan to evaluate these trials.

Slopes. We compared the average slopes of the best-fitting lines in the MI task with those in the II task to determine whether participants scanned iconic images and mental images at the same rate. The mean slope in the MI task ($M = 25$ msec/cm) was significantly steeper than the slope in the II task ($M = 12$ msec/cm) [$t(35) = 5.58, p < .0005$]. The results suggest that participants in this experiment scanned iconic images about twice as fast as they scanned mental images.

Intercepts. As reported in Experiment 1, the height of the intercept in the MI task ($M = 1,204$ msec) was greater than the height in the II task ($M = 409$ msec) [$t(35) = 11.75, p < .0005$]. This difference in the intercepts was expected because participants had to generate a mental image of the dots in the MI task, and this time is included in the height of the intercept.

ERs. We compared the ERs in the two scanning tasks on *yes* and *no* trials. Participants made fewer errors in the II task ($M = 5.3\%$) than in the MI task ($M = 9.7\%$) [$t(35) = 4.41, p < .0005$]. As opposed to what we found in Experiment 1, the MI task of Experiment 2 was more difficult than the II task.

Comparison of Experiments 1 and 2. The different slopes, intercepts, and the ERs in the II and MI tasks reflect differences in the MI tasks of Experiment 1 versus Experiment 2. We found steeper slopes in the MI task in Experiment 2 ($M = 25$ msec/cm) than in Experiment 1 ($M = 16$ msec/cm) [$t(70) = -2.68, p < .005$]; the intercept was higher in Experiment 2 ($M = 1,204$ msec) than in Experiment 1 ($M = 864$ msec) [$t(70) = -4.23, p < .0001$], and the ERs were larger in Experiment 2 ($M = 9.7\%$) than in Experiment 1 ($M = 6.4\%$) [$t(70) = -2.28, p < .025$]. One possible explanation of these differences is that when participants had to rely on their long-term memory of the dots (Experiment 2), the image was lower quality than it was when it was formed on the basis of a just-seen display (Experiment 1). And, because the image was degraded, the participants scanned it more slowly, required more time to generate it, and made more errors.

In contrast with what we found in the MI task, the mean slopes in the II task did not differ between Experiment 1 ($M = 14$ msec/cm) and Experiment 2 ($M = 12$ msec/cm) [$t(70) = 1.66, n.s.$], nor did the ERs ($M = 6.2\%$ in Experiment 1 and $M = 5.3\%$ in Experiment 2, $t < 1$). These results suggest that even when the pattern of dots and the arrow were presented simultaneously (Experiment 1), an iconic image was used to perform the task as in Experiment 2.

Correlational Analyses

The logic of the correlational analyses was identical to that in Experiment 1: We expect high correlations between the slopes if, in fact, the same scanning process is used in both tasks. We considered the same four measures as we did in Experiment 1 for each participant on each task and checked the split-half reliability for each of them. Critically, the slope measures all were reliable [$r(34) = .89, p < .01$, for MI, and $r(34) = .92, p < .01$, for II]. Reliability coefficients for the intercept and the ERs ranged from .95 to .99 (all $ps < .01$). And, unlike what we found in Experiment 1, the correlations between time and distances were reliable [$r(34) = .70, p < .01$, for MI, and $r(34) = .75, p < .01$, for II]. The correlations between all dependent variables are presented in Table 2.

Slopes. The slopes of the best-fitting lines in the II task were highly correlated with those in the MI task [$r(34) = .85, p < .01$]. As in Experiment 1, we corrected the correlation between the slopes for attenuation (Spearman, 1907) and found that the slopes in the MI and II tasks were correlated extremely highly [$r(34) = .94, p < .01$]. Finally, we used Fisher's (1921) z -transformation to compare the correlation coefficients between the scanning tasks in Experiments 1 and 2. The correlation (corrected by attenuation) between the slopes in the MI and II tasks in Experiment 2 was stronger than the correlation between the MI and II slopes in Experiment 1 ($z = -2.48, p < .01$). By restricting eye movements in the MI task (as opposed to the procedure in Experiment 1), it seems that the relationship between participants' scanning efficiency was strengthened. Not surprisingly, given these results, we found that the coefficients of correlation between time and distance in the MI and II tasks themselves were correlated [$r(34) = .61, p < .01$, and $r(34) = .84, p < .01$, when corrected by attenuation].

Table 2
Experiment 2: Matrix of Correlations

	Task	Slope		Intercept		Coefficient		Error Rate	
		MI	II	MI	II	MI	II	MI	II
Slope	MI	1.00	.85**	.20	.06	.56**	.54**	.00	-.02
	II		1.00	.00	.02	.62**	.65**	-.04	.02
Intercept	MI			1.00	.43**	-.41*	-.09	.19	.03
	II				1.00	-.10	.03	.20	.07
Coefficient	MI					1.00	.61**	-.17	-.09
	II						1.00	-.07	.04
Error rate	MI							1.00	.57**
	II								1.00

Note— Slope, slope of the best-fitting line; Intercept, intercept at the ordinate; Coefficient, coefficient of correlation between times and distances to scan; MI, the mental image scanning task; FV, the free visual scanning task; II, the iconic image scanning task. * $p < .05$. ** $p < .01$.

Intercepts. We also correlated the intercepts between the MI task and the II task [$r(34) = .43, p < .01$; when corrected by attenuation, the correlation was $r(34) = .45, p < .01$]. However, the intercepts and the slopes were not correlated, which suggests that the slopes and the intercepts reflect different aspects of processing. Finally, the comparison of the coefficient of correlations between the slopes in the MI and II tasks and the intercepts in those tasks (using Fisher's [1921] procedure) revealed that the slopes were more highly correlated than were the intercepts ($z = 5.14, p < .0001$). Taken together, the results indicate that the relationship between participants' efficiency in scanning a mental image and an iconic image is not driven by more general processes.

ERs. Error rates in the MI and II tasks were also correlated [$r(34) = .57, p < .01$; $r(34) = .58, p < .01$, when corrected by attenuation]. In contrast with Experiment 1, the ERs were not correlated with the slopes, intercepts, or with the coefficient of correlations (in each of the two scanning tasks), which rules out a speed-accuracy trade-off on all three measures. Most important for present purposes, the efficiency of the scanning process (revealed by the slopes) per se was not affected by a speed-accuracy trade-off.

Discussion

In the revised version of the mental image scanning task and the iconic image scanning task, as in Experiment 1, the time to scan increasing distances increased linearly as the distance between the tip of the arrows and the dots increased. In the II task, the average slopes of the best-fitting lines and the error rates were comparable to those we obtained in Experiment 1. Thus, it is unlikely that participants were engaged in a form of "memory scanning" during the II task of Experiment 1 (arising from the repeated presentation of the same pattern of dots); rather, it seems that participants scanned iconic images, as in Experiment 2 (when new stimuli were presented on each trial of a given block). Furthermore, the similarity of the slopes in the II tasks of Experiments 1 and 2 supports the inference that participants did not begin to scan the stimulus before it was removed in the iconic image scanning task of Experiment 1.

In contrast, in the MI task, not only the slopes but also the height of the intercepts and the accuracy of responses were affected by requiring the participants to generate the image on the basis of information stored in long-term memory, as opposed to on the basis of a just-seen display. The pattern of results revealed that when participants created the to-be-scanned image from long-term memory (Experiment 2), the task was more difficult than when the pattern was encoded immediately prior to generating the image (Experiment 1).

In addition, in Experiment 2, the participants scanned at different rates in the MI and the II tasks. Nevertheless, scanning abilities were strongly correlated between the iconic image scanning condition and the mental image scanning condition, which is consistent with the idea that the same processes were at play in both sorts of scanning. Moreover, the correlations in slopes do not reflect more general aspects of processing, as reflected by the lower

correlations between the intercepts and the fact that the intercepts were not correlated with the slopes.

GENERAL DISCUSSION

The two experiments reported in the present article converge in showing that participants who were better at scanning distances perceptually were also better at scanning distances across a mental image. The similarity of the slopes in Experiment 1 between the three scanning tasks led us to suggest that the underlying representations share structural properties. By the same token, the strong correlation of slopes between another version of the iconic and mental imagery scanning tasks in Experiment 2 is also consistent with the idea that the same type of representation was processed in the two tasks. In all cases, depictive representations apparently were processed, which accounts for the increased time to scan increased distances. In addition, the high correlations between scanning in the perceptual and mental imagery conditions are good evidence that the same type of representation was processed in the two cases—particularly in light of the weaker correlations with estimates of general processing speed.

The fact that the mental images were more difficult to scan in Experiment 2 than in Experiment 1 suggests an intriguing hypothesis: Perhaps mental images that are generated on the basis of information stored in long-term memory are akin to a degraded iconic image. If so, then it should be possible to degrade the iconic image in small steps and to find a version in which the slopes match those in the mental imagery condition—a way to "externalize" the quality of a mental image. Indeed, this approach could lead to a new, objective way to assess individual differences in the quality of mental images.

However, the differences in the slopes in Experiment 2 between the iconic image scanning task and the image scanning task introduce a note of caution. The steeper slopes observed in the mental imagery condition of Experiment 2 could simply represent a slower scanning process over the identical type of representation, which would make sense if the images generated from long-term memory are degraded (in comparison with those generated on the basis of perceptual input that was just encoded). Alternatively, these steeper slopes could indicate that the representation or processing is different in the two mental imagery conditions. Perhaps, for example, mental images generated from long-term memory are more fragmented than those generated on the basis of a just-seen stimulus, and part of the scanning process involves filling in missing material. If so, then image generation processes could play a more crucial role in the sort of scanning observed in Experiment 2 than the sort observed in Experiment 1. Nevertheless, the increase in time to scan increasing distance suggests that image representations depict information, regardless of the quality of the image.

The depictive representations that are scanned are not likely to be implemented in the topographically organized cortical areas 17 or 18, for several reasons. First, Kosslyn, Ball, and Reiser (1978) found that the time to scan increases at the same rate between two points that were

both “visible” in an image as it does between a point that initially was “visible” and a second one that initially was “offscreen” (as it were). Such findings suggest that a spatial representation elsewhere in the brain—such as in the posterior parietal cortex—may specify the spatial layout of scenes and be operated on during scanning. And, in fact, that area is activated during image scanning tasks (Ghaëm et al., 1997; Mellet et al., 2000; Mellet, Tzourio, Denis, & Mazoyer, 1995). That said, we also note that spatial representations involve other brain structures. For example, Epstein and Kanwisher (1998) presented evidence that a particular area within human parahippocampal cortex encodes the geometry of the local environment.

Second, areas 17 and 18 devote more cortex to the central regions of the visual field; hence, we should not expect a constant rate of scanning as distance increases (unless we assume that attentional processing shifts more rapidly over longer distances of cortex when the process first begins, which seems implausible). However, if image scanning relies on the mechanism proposed by Kosslyn and Schwartz (1977), in which the material is shifted across the visual buffer during scanning (which explains how people scan to part of the scenes initially “offscreen”), then these brain areas could be involved.

Third, portions of the posterior parietal lobe are also topographically organized (Serenó, Pitzalis, & Martinez, 2001); hence, they could play a role in depicting the spatial layout of imaged objects. This suggestion is consistent with the fact that eyeblinks—which typically disrupt iconic images—affect not only how long information persists in V1, but also activation in the posterior parietal cortex (Bodis-Wollner, Bucher, & Seelos, 1999; Hari, Salmelin, Tissari, Kajola, & Virsu, 1994); this result supports the idea that spatial iconic images rely on the posterior parietal cortex. Moreover, the posterior parietal lobe is also activated during the disengagement and the reengagement of visuospatial attention (see, e.g., Nobre et al., 1997; Posner, Walker, Friedrich, & Rafal, 1984, 1987), which are required during scanning.

The present findings support the claim that image representations depict information in the same way that visual representations do. We have particularly good evidence that perceptual and mental image representations are structurally similar when mental images are generated on the basis of recently encoded perceptual information. We know that many visual areas in the brain (at least 15 in the monkey brain; Felleman & Van Essen, 1991) depict information during perception. Thus, it follows that brain areas involved in depicting information during perception could be also be involved in mental imagery. By providing evidence that the structure of visual mental image representations is similar to that of visual perceptual representations, we provide additional evidence that the two representations specify information in the same way.

AUTHOR NOTE

This research was supported by the National Institutes of Health, Grant 2 R01 MH060734; any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and

do not necessarily reflect the views of NIH. We are grateful to Leah Jean McDonald, Lirui Li, and Csaba Orban for their help in constructing stimuli, recruiting participants, and collecting data. Correspondence concerning this article should be addressed to G. Borst, Harvard University, Department of Psychology, William James Hall 836, 33 Kirkland St., Cambridge, MA 02138 (e-mail: borst@wjh.harvard.edu).

REFERENCES

- AVONS, S. E., & PHILLIPS, W. A. (1980). Visualization and memorization as a function of display time and poststimulus processing time. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 407-420.
- BAHLL, A. T., & STARK, L. (1979). The trajectories of saccadic eye movements. *Scientific American*, *240*, 108-117.
- BASSO, A., BISIACH, E., & LUZZATTI, C. (1980). Loss of mental imagery: A case study. *Neuropsychologia*, *18*, 435-442.
- BECKER, M. W., PASHLER, H., & ANSTIS, S. M. (2000). The role of iconic memory in change-detection tasks. *Perception*, *29*, 273-286.
- BEECH, J. R. (1980). Imaginal vs. perceptual scanning of a visual representation. *Perceptual & Motor Skills*, *50*, 367-370.
- BEHRMANN, M., WINOCUR, G., & MOSCOVITCH, M. (1992). Dissociation between mental imagery and object recognition in a brain-damaged patient. *Nature*, *359*, 636-637.
- BISIACH, E., & LUZZATTI, C. (1978). Unilateral neglect of representational space. *Cortex*, *14*, 129-133.
- BODIS-WOLLNER, I., BUCHER, S. F., & SEELOS, K. C. (1999). Cortical activation patterns during voluntary blinks and voluntary saccades. *Neurology*, *53*, 1800-1805.
- BORST, G., KOSSLYN, S. M., & DENIS, M. (2006). Different cognitive processes in two image-scanning paradigms. *Memory & Cognition*, *34*, 475-490.
- 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.
- COLTHEART, M. (1983). Iconic memory. *Philosophical Transactions of the Royal Society B*, *302*, 283-294.
- CRAVER-LEMLEY, C., & REEVES, A. (1987). Visual imagery selectively reduces vernier acuity. *Perception*, *16*, 599-614.
- CRAVER-LEMLEY, C., & REEVES, A. (1992). How visual imagery interferes with vision. *Psychological Review*, *99*, 633-649.
- DENIS, M. (1991). *Image and cognition* (M. Denis & C. Greenbaum, Trans.). New York: Harvester.
- DENIS, M., & COUDE, M. (1989). Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology*, *1*, 293-307.
- DENIS, M., & KOSSLYN, S. M. (1999). Scanning visual mental images: A window on the mind. *Current Psychology of Cognition*, *18*, 409-465.
- DROR, I. E., & KOSSLYN, S. M. (1994). Mental imagery and aging. *Psychology & Aging*, *9*, 90-102.
- DUYSSENS, J., ORBAN, G. A., CREMIEUX, J., & MAES, H. (1985). Visual cortical correlates of visible persistence. *Vision Research*, *25*, 171-178.
- ENGEL, G. R. (1970). An investigation of visual responses to brief stereoscopic stimuli. *Quarterly Journal of Experimental Psychology*, *22*, 148-166.
- ENNS, J. T., & DI LOLLO, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*, 345-352.
- EPSTEIN, R., & KANWISHER, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598-601.
- FARAH, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition*, *18*, 245-272.
- FARAH, M. J., LEVINE, D. N., & CALVANO, R. (1988). A case study of mental imagery deficit. *Brain & Cognition*, *8*, 147-164.
- FELLEMAN, D. J., & VAN ESSEN, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1-47.
- FINKE, R. A. (1985). Theories relating mental imagery to perception. *Psychological Bulletin*, *98*, 236-259.
- FINKE, R. A., & PINKER, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *8*, 142-147.
- FINKE, R. A., & PINKER, S. (1983). Directional scanning of remembered

- visual patterns. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 398-410.
- FISHER, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3-32.
- FRANCIS, G. (1996). Cortical dynamics of visual persistence and temporal integration. *Perception & Psychophysics*, **58**, 1203-1212.
- FUCHS, A. F. (1976). The neurophysiology of saccades. In R. A. Monty & J. W. Senders (Eds.), *Eye movements and psychological processes* (pp. 39-53). Hillsdale, NJ: Erlbaum.
- GANIS, G., SCHENDAN, H. E., & KOSSLYN, S. M. (2007). Neuroimaging evidence for object model verification theory: Role of prefrontal control in visual object categorization. *NeuroImage*, **34**, 384-398.
- GANIS, G., THOMPSON, W. L., & KOSSLYN, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, **20**, 226-241.
- GEGENFÜRTNER, K. R., & SPERLING, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 845-866.
- GHAËM, O., MELLET, E., CRIVELLO, F., TZOURIO, N., MAZOYER, B., BERTHOZ, A., & DENIS, M. (1997). Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. *NeuroReport*, **8**, 739-744.
- GUARIGLIA, C., PADOVANI, A., PANTANO, P., & PIZZAMIGLIO, L. (1993). Unilateral neglect restricted to visual imagery. *Nature*, **364**, 235-237.
- HARI, R., SALMELIN, R., TISSARI, S. O., KAJOLA, M., & VIRSU, V. (1994). Visual stability during eyeblinks. *Nature*, **367**, 121-122.
- INTONS-PETERSON, M. J. (1983). Imagery paradigms: How vulnerable are they to experimenters' expectations? *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 394-412.
- ISHAI, A., UNGERLEIDER, L. G., MARTIN, A., & HAXBY, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, **12**, 35-51.
- JOLICŒUR, P., & KOSSLYN, S. M. (1985). Is time to scan visual images due to demand characteristics? *Memory & Cognition*, **13**, 320-332.
- KELLEY, T. A., SERENCES, J. T., GIESBRECHT, B., & YANTIS, S. (2008). Cortical mechanisms for shifting and holding visuospatial attention. *Cerebral Cortex*, **18**, 114-125.
- KOSSLYN, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics*, **14**, 90-94.
- KOSSLYN, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- KOSSLYN, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- KOSSLYN, S. M., BALL, T. M., & REISER, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 47-60.
- KOSSLYN, S. M., & SCHWARTZ, S. P. (1977). A simulation of visual imagery. *Cognitive Science*, **1**, 265-295.
- KOSSLYN, S. M., & THOMPSON, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, **129**, 723-746.
- 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., & GANIS, G. (2006). *The case for mental imagery*. New York: Oxford University 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.
- MAST, F. W., & KOSSLYN, S. M. (2002). Eye movements during visual mental imagery. *Trends in Cognitive Sciences*, **6**, 271-272.
- MELLET, E., BRISCOGNE, S., TZOURIO-MAZOYER, N., GHAËM, O., PETIT, L., ZAGO, L., ET AL. (2000). Neural correlates of topographic mental exploration: The impact of route versus survey perspective learning. *NeuroImage*, **12**, 588-600.
- 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.
- NOBRE, A. C., SEBESTYEN, G. N., GITELMAN, D. R., MESULAM, M. M., FRACKOWIAK, R. S., & FRITH, C. D. (1997). Functional localization of the system for visuospatial attention using positron emission tomography. *Brain*, **120**, 515-533.
- PAIVIO, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- PERKY, C. W. (1910). An experimental study of imagination. *American Journal of Psychology*, **21**, 422-452.
- PINKER, S. (1980). Mental imagery and the third dimension. *Journal of Experimental Psychology: General*, **109**, 354-371.
- PINKER, S., CHOATE, P. A., & FINKE, R. A. (1984). Mental extrapolation in patterns constructed from memory. *Memory & Cognition*, **12**, 207-218.
- POSNER, M. I., NISSEN, M. J., & OGDEN, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. L. Pick, Jr., & E. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 137-157). Hillsdale, NJ: Erlbaum.
- POSNER, M. I., WALKER, J. A., FRIEDRICH, F. J., & RAFAL, R. D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience*, **4**, 1863-1874.
- POSNER, M. I., WALKER, J. A., FRIEDRICH, F. J., & RAFAL, R. D. (1987). How do the parietal lobes direct covert attention? *Neuropsychologia*, **25**, 135-145.
- PYLYSHYN, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, **80**, 1-24.
- PYLYSHYN, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, **88**, 16-45.
- PYLYSHYN, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral & Brain Sciences*, **25**, 157-238.
- PYLYSHYN, Z. W. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences*, **7**, 113-118.
- RAYNER, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372-422.
- SEGAL, S. J., & FUSELLA, V. (1970). Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology*, **83**, 458-464.
- SERENO, M. I., PITZALIS, S., & MARTINEZ, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, **294**, 1350-1354.
- SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- SPEARMAN, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology*, **18**, 161-169.
- SPERLING, G. (1960). The information available in brief visual presentation. *Psychological Monographs: General & Applied*, **74**(11, Whole No. 498), 1-29.
- STEIGER, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, **87**, 245-251.
- SUPÈR, H., SPEKREIJSE, H., & LAMME, V. A. F. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science*, **293**, 120-124.
- TOOTELL, R. B., SILVERMAN, M. S., SWITKES, E., & DE VALOIS, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, **218**, 902-904.
- WEDE, J., & FRANCIS, G. (2006). The time course of visual afterimages: Data and theory. *Perception*, **35**, 1155-1170.
- WILMIZIG, C., SCHNEIDER, S., & SCHÖNER, G. (2006). The time course of saccadic decision making: Dynamic field theory. *Neural Networks*, **19**, 1059-1074.
- WILLIAMS, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B*, **21**, 396-399.