THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2010, 63 (10), 2031–2050

Psychology Press
Taylor & Francis Group

Individual differences in spatial mental imagery

Grégoire Borst and Stephen M. Kosslyn
Harvard University, Cambridge, MA, USA

In this article, we report a new image-scanning paradigm that allowed us to measure objectively individual differences in spatial mental imagery—specifically, imagery for location. Participants were asked to determine whether an arrow was pointing at a dot using a visual mental image of an array of dots. The degree of precision required to discriminate "yes" from "no" trials was varied. In Experiment 1, the time to scan increasing distances, as well as the number of errors, increased when greater precision was required to make a judgement. The results in Experiment 2 replicated those results while controlling for possible biases. When greater precision is required, the accuracy of the spatial image becomes increasingly important—and hence the effect of precision in the task reflects the accuracy of the image. In Experiment 3, this measure was shown to be related to scores on the Paper Folding test, on the Paper Form Board test, and on the visuospatial items on Raven's Advanced Progressive Matrices—but not to scores on questionnaires measuring object−based mental imagery. Thus, we provide evidence that classical standardized spatial tests rely on spatial mental imagery but not object mental imagery.

Keywords: Spatial mental imagery; Individual differences; Spatial ability; Image scanning.

Mental imagery has long held a unique place in psychology. It began as one of the central topics in experimental psychology, only to be banished from the field by the behaviourists. However, by the mid-1960s rigorous studies of the role of imagery in memory (e.g., Bower, 1970; Paivio, 1971) and of the relationship between imagery and perception (e.g., Segal & Fusella, 1969) rehabilitated this topic. Today, research on mental imagery is commonplace in experimental psychology and cognitive neuroscience (e.g., for a review see Kosslyn, Thompson, & Ganis, 2006).

Almost since its inception as a topic of scientific study, researchers have emphasized that people differ markedly in their imagery abilities (e.g., Galton, 1883; Marks, 1977)—but this aspect of imagery has yet to become a major focus of objective study. For the most part, subjective ratings are used to assess individual differences in imagery, and such ratings only sporadically predict performance in visuospatial tasks (e.g., Carpenter & Just, 1986; Kyllonen, 1996; Lohman, 1996; Mumaw, Pellegrino, Kail, & Carter, 1984; Pellegrino & Kail, 1982; Poltrock & Agnoli, 1986). For

Correspondence should be addressed to , Grégoire Borst, Harvard University, Department of Psychology, William James Hall 836, 33 Kirkland Street, Cambridge, MA, 02138, USA. E-mail: borst@wjh.harvard.edu

This material is based upon work supported by the National Science Foundation (NSF) under Grant REC-0411725; any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF. We are grateful to Katie Lewis, Jake Kantrowitz, Csaba Orban, Pooja Patnaick, and Magdalena Surma for their help in constructing stimuli, recruiting participants, and collecting data.

http://www.psypress.com/qjep
DOI:10.1080/17470211003802459

example, researchers have found little or no correlation between rated vividness of imagery (using the Vividness of Visual Imagery Questionnaire, VVIQ, Marks, 1973) and the performance on spatial abilities tests (e.g., Danaher & Thoresen, 1972; Di Vesta, Ingersoll, & Sunshine, 1971; Durndell & Wetherick, 1976a, 1976b; Ernest, 1977; Kosslyn, Brunn, Cave, & Wallach, 1984; Lorenz & Neisser, 1985; Paivio, 1971; Poltrock & Brown, 1984; Rehm, 1973; Richardson, 1977; Sheenan & Neisser, 1969).

Ratings of how vivid objects seem in mental images may not predict spatial abilities for a simple reason: Visual mental imagery is the product of a collection of different abilities (see Kosslyn et al., 2006), and such ratings tap only one such ability. Just as visual perception relies on separate systems that process properties of objects (such as shape and colour) and that process spatial properties (such as size and location), the same is true of imagery (Kosslyn, 1994; Levine, Warach, & Farah, 1985). In addition, individual differences in the two imagery abilities predict different types of performance. For example, Blajenkova, Kozhevnikov, and Motes (2006) developed a questionnaire (the Object Spatial Imagery Questionnaire, OSIQ) to measure individual differences in preferences and experiences in object and spatial mental imagery. Scores on the object versus spatial scales selectively correlated with scores on object versus spatial tests (e.g., Kozhevnikov, Kosslyn, & Shephard, 2005). Moreover, scientists tended to have higher scores on the spatial scales whereas visual artists had higher scores on the object scales. Dean and Morris (2003), using a different questionnaire, report consistent findings. Although intriguing, such questionnaires have the disadvantage of relying on self-report and not directly assessing imagery abilities.

To understand better the role of individual differences in mental imagery in cognitive tasks (such as in problem solving and learning), we need to develop objective measures of such individual differences. Moreover, we need to develop measures that tap specific imagery processes. In the experiments reported in this article, we focus on individual differences in spatial mental imagery per se. Spatial imagery consists of short-term spatial representations that are created on the basis of information stored in memory, not on the basis of immediate sensory input. We developed a new method to measure individual differences in the central aspect of spatial mental imagery—namely, imagery for spatial location. We use this new method to examine whether standardized spatial tests rely on spatial mental imagery. It is potentially important to gather such evidence, given that spatial imagery may play a role not only in many forms of cognition, but also in intelligence more generally (cf. Deary, 2000).

The method we developed relies on a scanning paradigm first introduced by Finke and Pinker (1982) and later refined by Borst, Kosslyn, and Denis (2006). In this paradigm, a pattern of dots is presented on the screen; the pattern is then removed and is replaced by an arrow. Participants are instructed to decide whether the arrow points at a location previously occupied by one of the dots. As the distance between the arrow and the target dot increases, the time to make the decision increases, consistent with the inference that participants scan their mental image of the array of dots. And in fact, many studies have reported a linear increase in response times with increasing distances scanned (e.g., Borst & Kosslyn, 2008; Borst et al., 2006; Denis & Cocude, 1989; Dror & Kosslyn, 1994; Finke & Pinker, 1982, 1983; Pinker, Choate, & Finke, 1984). The scanning effect (i.e., linear increase of response times with increasing distance) suggests that the way the representations are processed reflects the spatial structure of the representations used in these tasks (Kosslyn, 1972; Kosslyn, Ball, & Reiser, 1978; Pinker, 1980). Consequently, structural properties of the representation can be inferred from the behavioural data.

In the present studies, we used a modified version of the image-scanning paradigm of Borst and Kosslyn (2008) to assess individual differences in the precision of the spatial mental images generated from information stored in long-term memory. In Experiments 1 and 2 we modify this task to allow us to assess the precision of the spatial mental images. In Experiment 3, we provide evidence that standardized spatial tests and visuospatial items of Raven's Advanced Progressive Matrices rely on spatial mental imagery.

In Experiment 1, we systematically varied the degree of precision of the spatial information required to perform the task—namely, to decide whether an arrow pointed at a position previously occupied by one of the dots in a memorized array. If the behavioural data reflect the spatial structure of the underlying representation, then we should observe an effect of the degree of precision of the spatial information on the behavioural data (i.e., slopes of the best fitting lines and response accuracy). In Experiment 2, we controlled for a potential bias that could have produced the results observed in Experiment 1. Finally, in Experiment 3 we investigated whether individual differences in spatial ability as assessed by classical spatial tests (i.e., Paper Folding and Paper Form Board) are related to the ability to generate high-resolution spatial mental images from information stored in long-term memory as assessed by the behavioural measures taken in the modified images scanning task. If spatial tests do require spatial mental images, we expect a substantial correlation between behavioural data in the image-scanning task and scores on the spatial abilities tests.

EXPERIMENT 1

In this experiment, participants memorized a pattern of dots prior to the task. Following this, they visualized the dots in order to decide whether an arrow would have pointed at one of them, if the dots were on the screen as they had appeared when memorized. We varied how precisely the locations of the dots had to be specified in order to perform the task, and we defined the degree of precision of the spatial information as an *area of uncertainty*, hereafter referred as the AoU, surrounding each dot. As the radius of the AoU decreased, the location of the dots needed to be more accurate to perform the task. We used four radii of the AoU, which are hereafter referred to as different levels of AoU. The AoU affected only the way we designed the No arrows (those that did not point at a dot): For each level, the No arrows were aligned with one of the tangents of the AoU. We expected that as the radius of the AoU decreased, scanning would be slower (as reflected by the increase in time to scan greater distances—that is, the slope of response times (RTs) over distance) and accuracy would decrease.

Method

Participants

A total of 24 volunteers from Harvard University and the local community participated in this study (14 females and 10 males). All participants received pay or course credit. Their average age was 20 years, 8 months; 22 were right-handed, 2 left-handed. All reported normal or corrected-to-normal vision. Data from 3 additional people were not analysed because they performed the task at chance levels of performance. All 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 designed one configuration of four black dots, with each dot being 7 mm in diameter (subtending 0.5° visual angle). The configurations were placed in a 19-cm × 19-cm (14.4° × 14.4° visual angle) white square that was surrounded by a black frame. We created one set of 96 arrows for each of the four radii of the AoU. Arrows were 2 cm long (1.7° visual angle); 48 of them pointed directly at the centre of one of the dots (Yes arrows), and 48 missed all the dots (No arrows). Each arrow was placed at one of four possible distances from the target dot (or nearest dot, for the No arrows), ranging from 3 cm to 7.5 cm, with 1.5-cm increments of differences in distance. For each dot, 12 arrows pointed at it, 3 at each of the four distances. The four radii of the AoU ranged from 16.5 mm to 28.5 mm (hereafter referred as Level 1−Level 4), with 4-mm increments of differences in radius; each of the 48 No arrows was aligned with one

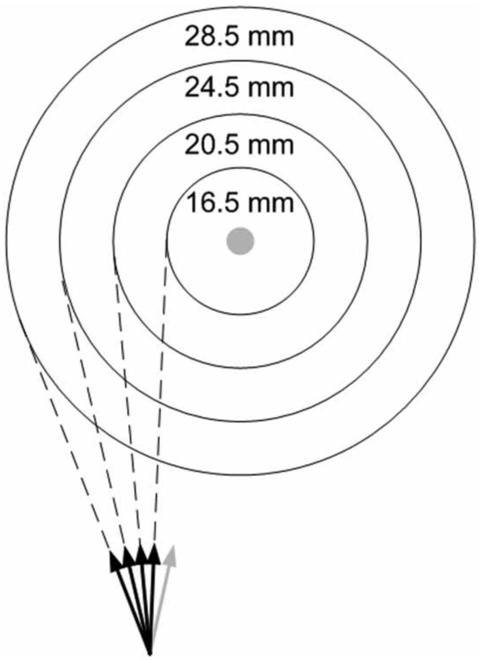

Figure 1. *Experiment 1: Principles of construction of the areas of uncertainty (AoUs). The location of the dot is represented in grey. The grey arrow represents an arrow that points at a dot. Black arrows represent arrows that miss the dots. Each circle represents a different radius of the AoU. Dashed lines represent the tangent of each circle on which a No arrow is aligned. Note that the angle with which black arrows miss the dots increases as the AoU radius increases.*

of the tangents of the AoU surrounding each dot at one of the four possible distances (see Figure 1). Consequently, as the radius of the AoU increased, the angle (averaged over the four distances) with which the No arrows missed the dots increased (respectively, $M = 21^\circ$ for a radius of 16.5 mm, $M = 27^\circ$ for a radius of 20.5 mm, $M = 33^\circ$ for a radius of 24.5 mm, and $M = 40^\circ$ for a radius of 28.5 mm). None of the arrows was strictly horizontal or vertical because we were concerned that scanning along the horizontal or the vertical axis could be different from scanning in other orientations. Each arrow was tilted at one of four possible angle ranges relative to the horizontal axis ($15^\circ{-}25^\circ$, $35^\circ{-}45^\circ$, $55^\circ{-}65^\circ$, or $75^\circ{-}85^\circ$). All arrows and dots were placed within a virtual circle with a 9-cm radius to prevent the participants from using the black frame as a reference for memorizing the positions of the dots.

In addition, the Yes arrows were designed to preclude "perceptual crowding" (Pylyshyn, 2002); as the distance increased between the tip of the arrow and the target dot, alternative dots did not become more crowded. The correlation between the distances and the angles of disparity between the direction of the arrows and the nearest alternative target dots across the four AoU levels was $r(190) = .11$, *ns*. Thus, "perceptual crowding" could not account for an increase in RTs with increasing distance.

A 17-inch monitor with resolution of 1,280 × 1,024 pixels and a refresh rate of 75 Hz was used to display the stimuli.

Procedure

The participants were tested individually, sitting approximately 75 cm from a computer screen. First, we asked the participants to study a pattern of dots on a hard-copy printout and then to draw the locations of the dots from memory on a blank sheet of paper, with both the 19-cm × 19-cm black frame and the fixation point printed on it. The black frame and the fixation point were identical to the ones displayed on the computer screen (same position and size) to ensure that the scale of the computer screen mapped onto the scale of the studied drawings. We printed a hard copy of the original patterns on transparency sheets. Participants superimposed the appropriate transparency on their drawing to compare their drawing to the original pattern. They were to note the disparities between the two and prepare to correct their mental image. They then redrew the drawing and again compared it to the original pattern. This draw-and-study procedure was repeated until all dots were drawn within 0.30 cm of their actual location two times in a row. On average, participants required 10 drawings to memorize the pattern of dots.

Following learning, participants were asked to follow the written instructions displayed on the screen. As part of these instructions, we showed the participants an example of a Yes trial in which the pattern of dots and an arrow that pointed at one of them were shown simultaneously. We explained that on each test trial a fixation cross would be displayed in the middle of the screen (which remained visible for 2.5 s). The participants were to visualize the dots at their exact locations, in the same form as they studied them, while keeping their eyes on the fixation cross. Following this, an arrow was displayed at an unexpected location on the screen in the black frame and remained visible until the participants responded (see Figure 2). We asked the participants to decide, as quickly and accurately as possible, whether the arrow pointed at a location occupied by one of the dots they memorized. Participants used their dominant hand to respond, pressing the "b" key if the arrow pointed at a location occupied by a dot they visualized and pressing the "n" key if not.

Each participant performed four separate blocks of trials, one for each level of AoU. The order of levels was fully counterbalanced over participants. Before each block, participants performed 24 practice trials where the computer provided feedback, which allowed them to become familiar with the No arrows at a specific level of AoU. We did not tell the participants about the radius of the AoU or that the radius of the AoU was varied, nor did we tell them that scanning was required to perform the task. The order of the trials within each block 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. The response times (RTs) and the nature of the response were recorded.

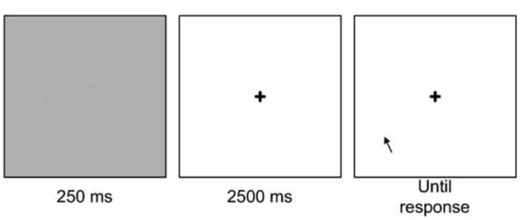

Figure 2. *Experiment 1: The procedure used in the image-scanning task.*

We also asked the participants to maintain their gaze on a fixation point for the complete duration of a given trial, to limit effects of eye movements on the scanning rate (see Bahill & Stark, 1979; Fuchs, 1976). 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. Finally, we asked the participants to complete a debriefing questionnaire at the end of the experiment, to ensure that they did not infer the purpose of the experiment and that they followed the instructions at least 75% of the time.

Results

As a first step, we analysed RTs to determine whether we replicated earlier findings of studies that used a scanning paradigm (i.e., a linear increase in RTs with increasing distance scanned). Following this, we compared the steepness of the slopes of the best fitting lines (i.e., scanning rates) and the error rates (ERs), to observe whether AoU had an effect on participants' scanning efficiency. In addition, for each of the analyses, we report the effect size of the analysis of variance (ANOVA; partial eta squared) or of the difference of the means (Cohen's *d*).

Preliminary analyses did not reveal any effect of gender on the RTs, the ERs, or the steepness of the slopes. Thus, we pooled the data over this variable, and we do not address it in the following report of the results.

Analysis of RTs and ERs

We analysed separately the RTs from correct responses on the Yes and No trials for each level of AoU. We expected participants to scan the entire distance in the Yes trials. For the No trials, when the radius of the AoU was large enough, it may have not been necessary for the participants to scan the entire distance from the tip of the arrow to the location of the dots because it was obvious along the way that the arrow missed the locations of the dots. In such cases, we cannot know at which point participants stopped scanning, and thus we did not expect distance to have an effect on the RTs. In addition, outliers were not included in the analysis; we defined outliers as either RTs greater than 2 standard deviations from the mean of that distance for that participant or RTs under 250 ms (because these RTs clearly did not reflect the time taken to scan). Outliers occurred on 3.6% of the trials.

Yes trials. We analysed separately the data for each of the four levels of AoU, to discover whether participants scanned a spatial mental image at each level. We averaged the RTs over trials for each distance for each participant. One-way repeated measures ANOVAs revealed significant effects of the distance on the RTs for each of the four AoU levels: for Level 1, $F(3, 69) = 24.94$, $MSE = 19{,}954.58$, $p < .0001$, $\eta_p^2 = .52$; for Level 2, $F(3, 69) = 11.70$, $MSE = 20{,}556.01$, $p < .0001$, $\eta_p^2 = .34$; for Level 3, $F(3, 69) = 15.67$, $MSE = 14{,}826.46$, $p < .0001$, $\eta_p^2 = .41$; and for Level 4, $F(3, 69) = 3.20$, $MSE = 16{,}232.45$, $p < .05$, $\eta_p^2 = .12$. In addition, as shown in Figure 3, the method of least squares revealed that RTs increased linearly with increasing distance, with $F(1, 23) = 88.39$, $MSE = 15{,}379.37$, $p < .0001$, $\eta_p^2 = .79$, for Level 1; $F(1, 23) = 30.41$, $MSE = 19{,}940.24$, $p < .0001$, $\eta_p^2 = .57$, for Level 2; $F(1, 23) = 57.60$, $MSE = 6{,}057.33$, $p < .0001$, $\eta_p^2 = .72$, for Level 3; $F(1, 23) = 46.98$, $MSE = 2{,}713.97$, $p < .0001$, $\eta_p^2 = .67$, for Level 4. In addition, for each of the four AoU levels, we computed the best fitting linear functions calculated by the method of least squares. RT and distance were highly correlated (with Bravais-Pearson *r*s ranging from .90 to .95, $p < .10$ in all cases). The results replicated earlier findings (Borst & Kosslyn, 2008; Finke & Pinker, 1982, 1983) and suggest that participants created a mental image of the pattern of dots and then scanned the distance between the tips of the arrows to the dots to make their decision.

No trials. Because each of the No arrows was positioned at one of the four possible distances from the nearest target dot, we analysed the RTs from the No trials in the same way that we analysed the RTs from the Yes trials. Repeated measures

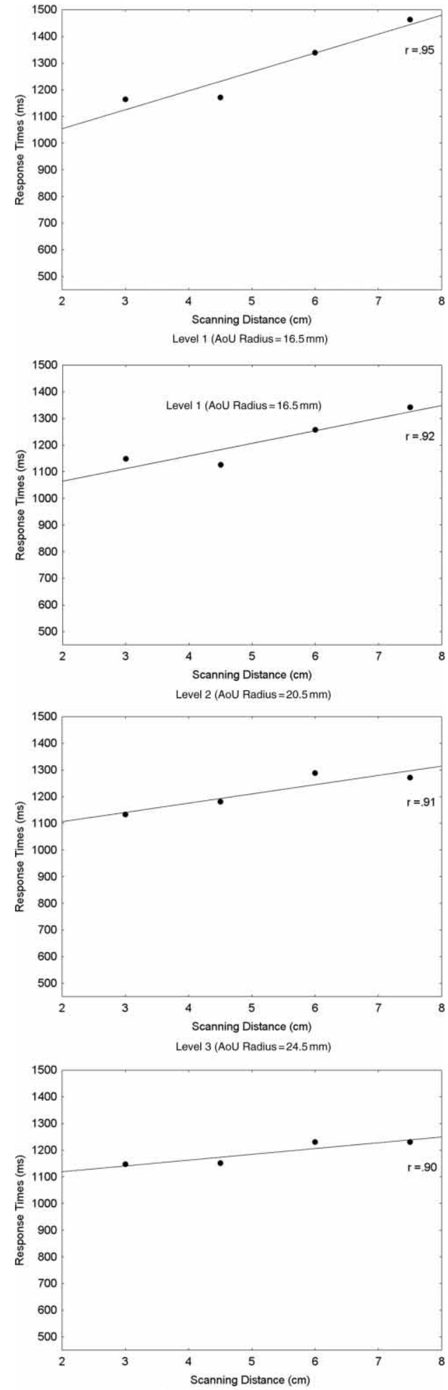

Figure 3. *Experiment 1: The time to scan increasing distances for different levels of the area of uncertainty (AoU).*

ANOVAs revealed a significant effect of distance on the RTs for each of the four levels of AoU, with $F(3, 69) = 19.99$, $MSE = 34{,}329.92$, $p < .0001$, $\eta_p^2 = .47$, for Level 1; $F(3, 69) = 8.31$, $MSE = 31{,}104.74$, $p < .0001$, $\eta_p^2 = .27$, for Level 2; $F(3, 69) = 7.70$, $MSE = 16{,}557.04$, $p < .0005$, $\eta_p^2 = .25$, for Level 3; and $F(3, 69) = 14.52$, $MSE = 12{,}557.08$, $p < .0001$, $\eta_p^2 = .39$, for Level 4. Moreover, the method of least squares revealed that RTs increased linearly as the distance between the No arrows and the nearest target dot increased; for Level 1, $F(1, 23) = 37.69$, $MSE = 54{,}391.22$, $p < .0001$, $\eta_p^2 = .62$; for Level 2, $F(1, 23) = 29.54$, $MSE = 24{,}911.79$, $p < .0001$, $\eta_p^2 = .56$; for Level 3, $F(1, 23) = 23.01$, $MSE = 16{,}338.50$, $p < .0001$, $\eta_p^2 = .50$; for Level 4, $F(1, 23) = 22.26$, $MSE = 10{,}812.13$, $p < .0001$, $\eta_p^2 = .49$.

Finally, RTs were highly correlated with distance only for the three first levels of AoU with the smaller radii (with *r*s ranging from .97 to .99, $p < .05$ in all cases). RTs were not significantly correlated with distance in the condition with the largest AoU, $r(2) = .67$, *ns*. As the radius of the AoU decreased, participants scanned in the direction indicated by the arrow up to the region of the nearest dots, as Finke and Pinker (1982, 1983) demonstrated. However, with the largest AoU (i.e., with the largest radius), the discrimination was easy enough that participants did not necessarily need to scan the entire distance to make their decision (as reported by Borst & Kosslyn, 2008).

Slopes. A one-way repeated measures ANOVA revealed that the slopes differed for different AoUs, $F(3, 69) = 22.68$, $MSE = 465.72$, $p < .0001$, $\eta_p^2 = .50$. The average slope was steeper at Level 1 (71 ms/cm) than at Level 2 (47 ms/cm), $t(23) = 3.22$, $p < .005$, $d = 0.61$; steeper at Level 2 than at Level 3 ($M = 35$ ms/cm), $t(23) = 2.23$, $p < .025$, $d = 0.40$; and steeper at Level 3 than at Level 4 ($M = 22$ ms/cm), $t(23) = 3.21$, $p < .005$, $d = 0.68$ (see Table 1). Thus, participants typically slowed their rate of scanning when the task required having a more precise representation of the position of the dots (i.e., when the radius of the AoU decreased).

Table 1. *Experiment 1: Mean slopes and error rates for different levels of area of uncertainty*

	AoU levels			
	Level 1	*Level 2*	*Level 3*	*Level 4*
Slopes	71 (7.6)	47 (8.6)	35 (7.1)	22 (3.2)
ER	29.6 (1.8)	27.6 (2.1)	24 (2)	20.5 (2.1)

Note: ER = error rate. AoU = area of uncertainty. Standard errors of the mean in parentheses.

ERs. As shown in Table 1, ERs (including both Yes and No trials) increased as AoU decreased, $F(3, 69) = 11.95$, $MSE = 33.50$, $p < .0001$, $\eta_p^2 = .35$. On the No trials, AoU affected ERs, $F(3, 69) = 78.1$, $MSE = 51.38$, $p < .0001$, $\eta_p^2 = .51$. Specifically, participants made more errors at Level 1 ($M = 32.6\%$) than at Level 2 ($M = 26.9\%$), $t(23) = 2.83$, $p < .005$, $d = 0.43$; more errors at Level 2 than at Level 3 ($M = 20.1\%$), $t(23) = 2.70$, $p < .01$, $d = 0.52$; and more errors at Level 3 than at Level 4 ($M = 16.7\%$), $t(23) = 2.11$, $p < .025$, $d = 0.29$. The effect of AoU on the ERs of the Yes trials was not significant, $F(3, 69) = 1.15$, $MSE = 64.21$, $p = .34$, and none of the specific comparisons of the ERs between adjacent levels of AoU were significant. Thus, the effects of AoU on the slopes of the RTs could not be attributed to a speed/accuracy trade-off.

Discussion

The fact that RTs increased as the distance to scan increased suggests that depictive representations were processed. This finding was a prerequisite for drawing inferences about the effects of the AoU on the precision of spatial mental images. The logic of our approach to measuring individual differences in spatial imagery rests on the idea that points in the image are represented in a coordinate space, such that we could estimate the spatial error surrounding each point. And in fact, we found that AoU affected both the rate of scanning (as revealed by the analysis of the steepness of the slopes) and accuracy. Thus, given the effect of the AoU on our behavioural measures, we have the first hint that both slopes and accuracy in the image-scanning task are valid measures of participants' ability to generate precise spatial mental images.

However, several aspects of the task could have biased the results and could have led us to overestimate the effect of the AoU on the dependent variables of the image-scanning task. Experiment 2 addresses the key issues.

EXPERIMENT 2

One could argue that our task does not tap the underlying spatial structure of the representation scanned. If so, we would not be justified in using this task to assess individual differences in the precision of spatial imagery. Four claims could be made to undermine our inferences: First, the No arrows were positioned at the same four possible distances as the Yes arrows; thus when the radius of the AoU decreased, the angle with which No arrows missed the dots decreased as well. Consequently, participants' scanning rates could have been slower for the smaller AoU levels not only because of the radius length of the AoU but also because of a "discrimination effect" (i.e., the angle with which No arrows were missing the dots). Second, within each level of the AoU, the farther away the No arrows were positioned from the nearest possible target dot, the smaller was the angle with which No arrows missed the dots. Thus, one could argue that the increased RTs with greater distance for the Yes trials was an indirect result of the "discrimination effect" on the No arrows. If so, then the slopes of the best fitting lines did not reflect positional uncertainty in the representation. Third, participants scanned the same pattern of dots for all four levels of AoU, which could have affected the differences in slopes among the levels of AoU.

In Experiment 2, we revised the procedure in the scanning task to address these issues. First, for all levels of AoU we kept constant the angle with which the No arrows missed the dots, which allowed us to determine whether the level of AoU alone was the factor responsible for the findings in Experiment 1. Consequently, the No arrows were positioned closer to the nearest possible target dot as the radius of the AoU became smaller. Second, within each level, because all No arrows were positioned at the same distance from the nearest possible target dot, if a linear increase in RTs with distance was found, it could not be attributed to a "discrimination effect". Third, participants memorized a new pattern of dots for each of the levels of AoU. Finally, we used the results of Experiment 1 to guide us in more precisely defining the levels of AoU, and we now included only three levels but with a larger increment in radius (6 mm as opposed to 4 mm).

Method

Participants

A total of 18 volunteers from Harvard University and the local community participated in this study (10 females and 8 males). All participants received pay or course credit. Their average age was 19 years, 8 months; 15 were right-handed and 3 left-handed. All reported normal or corrected-to-normal vision. Data from 2 additional people were not analysed because they performed the task at chance levels of performance. No participant had taken part in Experiment 1. All 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 addition to the pattern used in Experiment 1, we created two new patterns. Pattern 2 and Pattern 3 were created by rotating the pattern used in Experiment 1 by 90 degrees and by 180 degrees, respectively. We created three new sets of 96 arrows (48 Yes arrows and 48 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"). None of the arrows were strictly horizontal or vertical, and all arrows were placed within a virtual circle with a 9-cm radius.

The three radii of the AoU used in this experiment ranged from 14.5 mm to 26.5 mm, with 6-mm increments of differences in radius (hereafter referred as Level 1–Level 3). Because we designed the No arrows to miss the dots by 20° at all three levels of AoU, the distance (d) between the tip of the No arrows and the nearest possible target dot increased as the radius of the AoU increased (respectively, d = 40 mm for Level 1; d = 56 mm for Level 2; and d = 73 mm for Level 3, see Figure 4). Thus, this design precluded a "discrimination effect" for the No trials (i.e., an increase of the angle with which No arrow missed the dots when the radius of the AoU increased).

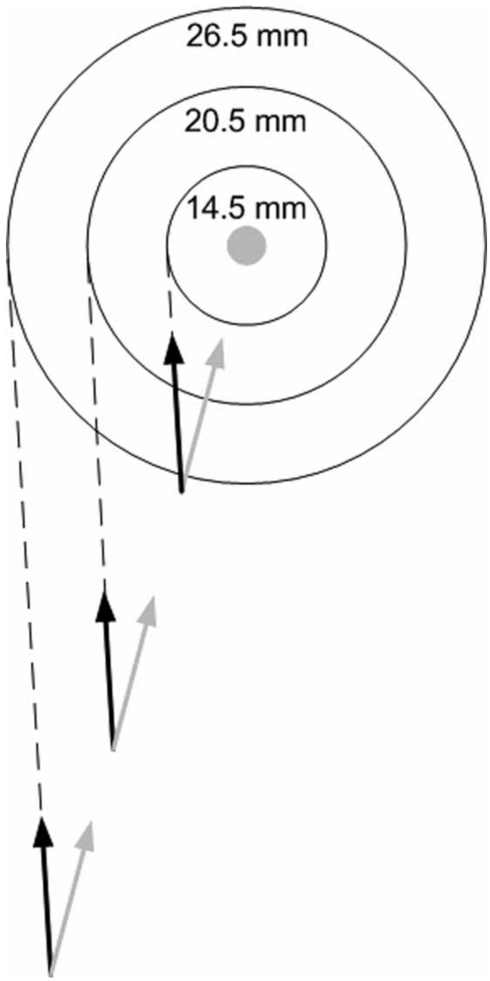

Figure 4. *Experiment 2: Principles of construction of the areas of uncertainty (AoU). The location of the dot is represented in grey. The grey arrows represent arrows that point at a dot. Black arrows represent the arrows that miss the dots. Each circle represents a different radius of the AoU. Dashed lines represent the tangent of each circle on which a No arrow is aligned. Note that the angle with which black arrows miss the dot is kept constant.*

We presented the stimuli on the same computer screen as that used in Experiment 1, with the same brightness and contrast settings.

Procedure

The procedure on each experimental trial was identical to that used in Experiment 1. We counterbalanced the pattern used for each level of AoU and the order of the three levels of the AoU between participants. As in Experiment 1, a draw-and-study procedure was used to ensure that participants memorized the dots. In this experiment, however, the participants memorized a new pattern immediately prior to each block of trials. On average, participants required 10 drawings to memorize the patterns of dots at each of the three levels of AoU, and there was no difference in the number of drawings necessary to reach criterion for each configuration, $F(2, 34) < 1$, *ns*.

Results

We analysed RTs and ERs in the same way as in Experiment 1. Preliminary analyses did not reveal an effect of gender, the order of the blocks, or the stimulus patterns on both dependent variables. Thus, we pooled the data over these variables, and we do not address them in the following report of the results.

Analysis of RTs and ERs

We conducted the same analyses as those used in Experiment 1. Defining outliers as in Experiment 1, 2.9% of the trials were considered as outliers.

Yes trials. One-way repeated measures ANOVAs revealed that RTs differed depending on the distance from the tip of an arrow to a location of a dot previously memorized, with $F(3,\ 51) = 32.37$, $MSE = 24{,}416.99$, $p < .0001$, $\eta_p^2 = .66$, for Level 1; $F(3,\ 51) = 9.19$, $MSE = 25{,}329.06$, $p < .0001$, $\eta_p^2 = .35$, for Level 2; and $F(2,\ 51) = 3.68$, $MSE = 18{,}061.97$, $p < .05$, $\eta_p^2 = .18$, for Level 3. In addition, as shown in Figure 5, RTs increased linearly with distance, as documented by the best fitting functions calculated by the method of least squares, with $F(1,\ 17) = 79.21$, $MSE = 29{,}866.87$, $p < .0001$, $\eta_p^2 = .82$, for Level 1; $F(1,\ 17) = 59.05$, $MSE = 10{,}839.61$, $p < .0001$, $\eta_p^2 = .78$, for Level 2; and $F(1,\ 17) = 38.49$, $MSE = 4{,}986.61$, $p < .0005$, $\eta_p^2 = .69$, for Level 3. Distance and RT were highly correlated (with *r*s ranging from .96 to .99, $p < .05$ in all cases). The data suggest that participants created a spatial mental image of the pattern of dots and scanned the distance between the tip of the arrows and the dot to decide whether the arrows pointed at one of the dots.

No trials. Because we kept constant the distance to the nearest possible target dot for each level of AoU, we could not analyse the data the same way that we did in Experiment 1. However, we compared the RTs between the different levels of AoU. A repeated measures ANOVA showed that RTs differed for the different AoUs, with $F(2,\ 34) = 14.13$, $MSE = 15{,}937.82$, $p < .0001$, $\eta_p^2 = .45$. Participants were faster for Level 3 ($M = 1{,}196$ ms) than for Level 2 ($M = 1{,}318$ ms), $t(17) = 2.73$, $p < .01$, $d = 0.42$ and were faster for Level 2 than for Level 1 ($M = 1{,}401$ ms), $t(17) = 2.38$, $p < .025$, $d = 0.29$. Thus, as the radius of the AoU decreased (and consequently the more precise the spatial mental image needed to be), the longer participants took to respond.

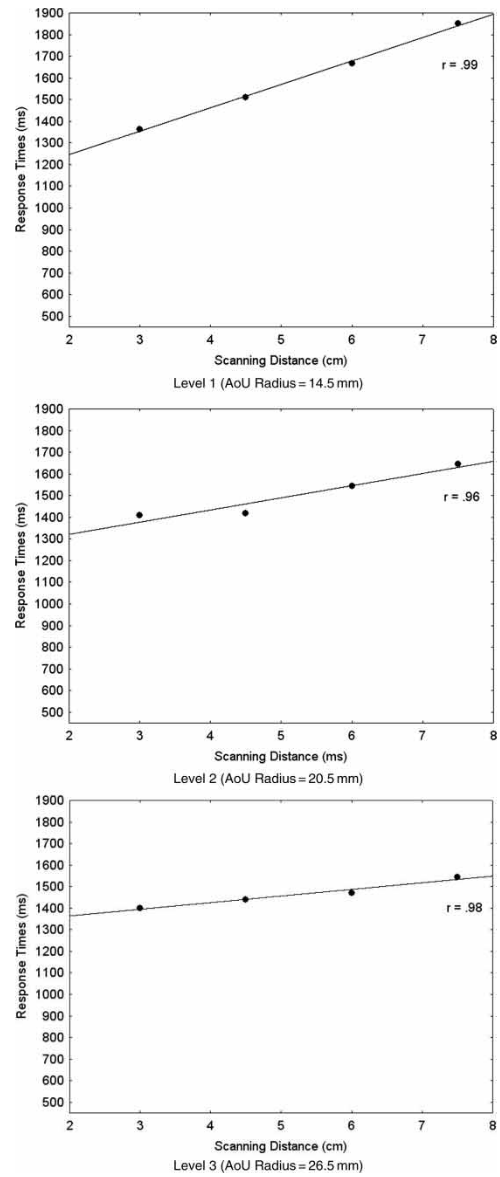

Figure 5. *Experiment 2: The time to scan increasing distances for different levels of the area of uncertainty (AoU).*

Slopes. The confounding between the size of the radius of the AoU and distance of the arrows (which was necessary to avoid discrimination effect) prevents us from using the RTs in No trials as evidence that smaller AoUs require more precise images. However, if in fact smaller AoUs require more precise images, then we expect the participants to scan more carefully, as reflected in the increase in time to scan greater distances (i.e., the slope of RT over distance). And in fact, the average slopes were affected by the levels of the AoU, $F(2,\ 34) = 62.99$, $MSE = 443.54$, $p <$

Table 2. *Experiment 2: Mean slopes and error rates for different levels of area of uncertainty*

	AoU levels		
	Level 1	*Level 2*	*Level 3*
Slopes	108 (12.1)	56 (7.3)	31 (4.9)
ER	33.4 (2.3)	28.5 (1.8)	22.6 (1.8)

Note: ER = error rate. AoU = area of uncertainty. Standard errors of the mean in parentheses.

.0001, $\eta_p^2 = .79$ (see Table 2). Participants scanned at a slower rate at the smallest level, Level 1 ($M = 108$ ms/cm), than at Level 2 ($M = 56$ ms/cm), $t(17) = 8.08$, $p < .0001$, $d = 1.23$, and they scanned at a slower rate at Level 2 than at Level 3 ($M = 31$ ms/cm), $t(17) = 6.02$, $p < .0001$, $d = 0.94$.

ERs. A repeated measures ANOVA revealed a significant effect of AoU on the ERs, $F(2, 34) = 17.21$, $MSE = 30.40$, $p < .0001$, $\eta_p^2 = .51$ (see Table 2). Participants made more errors for Level 1 ($M = 33.4\%$) than for Level 2 ($M = 28.5\%$), $t(17) = 2.69$, $p < .01$, $d = 0.57$, and they made more errors for Level 2 than at Level 3 ($M = 22.6\%$), $t(17) = 3.81$, $p < .005$, $d = 0.79$. On the No trials, AoU affected the ERs, $F(2, 34) = 24.92$, $MSE = 46.08$, $p < .0001$, $\eta_p^2 = .59$. Moreover, participants made more errors for Level 1 ($M = 31.6\%$) than for Level 2 ($M = 22.5\%$), $t(17) = 3.68$, $p < .005$, $d = 0.76$, and made more errors for Level 2 than for Level 3 ($M = 14.6\%$), $t(17) = 4.14$, $p < .0005$, $d = 0.88$. Finally, AoU did not affect the ERs in the Yes trials, $F(2, 34) = 2.91$, $MSE = 55.24$, $p = .07$. Thus, the increase of the scanning rate as the radius of the AoU increased could not be ascribed to a speed/accuracy trade-off, given that participants made more errors for the level of AoU where they scanned at a slower rate.

Discussion

As in Experiment 1, the time to scan increased linearly as the distance between the tip of the arrows and the dots increased. The fact that we observed this effect is important because—as opposed to Experiment 1—a "discrimination effect" could not occur on the No trials. Thus, it is of interest that we replicated and extended the findings of Experiment 1. First, we again found that AoU affected both the rate of scanning and accuracy. Second, by keeping constant the angle with which the No arrows missed the dots over the different levels of AoU, we demonstrated that the different scanning rates at the different levels should be attributed to the variation of the radii of the AoU per se. Finally, given that we provided evidence that a speed/accuracy trade-off could not account for the difference in scanning rates, we are confident that this paradigm offers a window on the structure of the representations scanned.

Moreover, in this study we observed individual differences (as evident in the standard errors of the means, shown in Table 2) both on the slopes and on the ERs—and hence the scanning task may be able to be used as an objective way to assess the quality of a person's spatial mental imagery. That is, we can determine the quality of the coordinate locations incorporated in a particular person's spatial representations by observing the effect of the variation of the AoU on the slope of the scanning function and on the number of errors. The logic is that as the radius of the AoU decreases, participants with the least accurate representation of the locations of the dots should make increasingly greater numbers of errors. In keeping with this logic, in Experiment 3 we attempt to demonstrate that spatial tests rely on spatial mental imagery.

EXPERIMENT 3

In this experiment, we investigated whether individual differences in performance in the image-scanning task are related to individual differences in performance on standard tests of spatial ability. If so, this would provide evidence that spatial mental images are used when one performs standardized spatial tests. In order to demonstrate that spatial tests rely on spatial mental imagery, but not on object based-mental imagery, we also asked whether scores on object-based imagery tasks are correlated with scores on spatial tests.

Thus, we administered an adaptation of the image-scanning task of Experiment 2, as well as: (a) the Paper Folding test (hereafter referred as PF test); (b) the Paper Form Board test (hereafter referred as PFB test); (c) the Raven's Advanced Progressive Matrices (hereafter referred as APM); (d) the OSIQ; and (e) the VVIQ. The two spatial tests were chosen from the kit of factor referenced tests because they loaded on a single visualization factor (Ekstrom, French, Harman, & Dermen, 1976). We also estimated *g* (i.e., general intelligence) for each participant through their score on the APM to demonstrate that if any relation arises between the scanning task and spatial performance, this relation is independent of the level of intelligence per se. In addition, we administered two mental imagery questionnaires (i.e., VVIQ and OSIQ), which have scales that assess properties of images of objects, because studies have shown that spatial and object representations are processed in different parts of the brain (e.g., Kosslyn, Thompson, Gitelman, & Alpert, 1998; Mishkin, Ungerleider, & Macko, 1983; Ungerleider & Mishkin, 1982). If people perform standardized spatial tests by using spatial mental images, but not object mental images, then performance on the image-scanning task should be correlated with scores on the spatial tests but not with scores on the object-imagery scale of the OSIQ or with the VVIQ.

Method

Participants

We tested 48 volunteers (26 females and 22 males) from Harvard University and the local community, with an average age of 23 years and 10 months and with normal or corrected-to-normal vision. A total of 44 participants were right-handed and 4 left-handed. All participants received pay or course credit, and none had participated in Experiments 1 or 2. Data from three additional people were not included because they performed the image-scanning task at a chance level of performance. All 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 and procedure

The participants were tested individually, sitting approximately 75 cm from the same computer screen used in Experiments 1 and 2.

All participants performed the different tasks in the same order: the image-scanning task, the PF test, the PFB test, the APM, the OSIQ, and the VVIQ. We asked the participants to follow the written instructions displayed on the screen or in the booklets that accompanied the paper-and-pencil tests and questionnaires. A 5-minute break was provided between each task. At the end of the experiment, participants completed a debriefing questionnaire to ensure that they were not aware of the hypotheses and that they followed the instructions at least 75% of the time.

Image-scanning task. In order to shorten the task, we used only Level 1 and Level 3 of the AoUs of Experiment 2. For each level, a different pattern of four dots was used (i.e., Pattern 1 of Experiment 2 for the trials at Level 1 and Pattern 3 of Experiment 2 for the trials at Level 3). All participants first performed a block of 96 Level 3 trials (the easy condition) and a block of 96 Level 1 trials (the difficult condition). The experimental trials were structured exactly the same as those in the previous experiments. RTs and the nature of the response were recorded. Before each block of experimental trials participants performed 24 practice trials where they received feedback on their responses.

Before performing each level of the image-scanning task, participants took part in the draw-and-study procedure used in Experiments 1 and 2, which was repeated until all four dots were drawn within 0.30 cm of their actual locations two times in a row. For each pattern of dots, participants required between 3 and 11 drawings to reach this criterion. The participants required a comparable number of drawings to learn the two configurations, $t(47) = 1.05$, $p > .25$.

PF test. In this test, a figure is presented that represents a square piece of paper that has been folded, with one or two circles drawn on it to show where holes were punched in the paper. To the right of this drawing, five figures are presented that show the positions of the holes when the paper is completely unfolded. Participants select which of the five figures would correspond to the unfolded version of the standard on the left. The PF has two parts, each of which has 10 items. Participants were given three minutes for each part. The score is the number of correct responses.

PFB test. In this test, a geometrical figure is presented on top of a page. The participants decide which pieces beneath each figure (from two to five) will make the complete figure when put together. The PFB consists of two parts, each of which has 24 items. Participants were given eight minutes for each part. The score is the number of items correctly answered.

APM test. We also gave the participants the Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998). This test consists of 36 items (Set II of the APM). The items are presented in increasing difficulty. For each item, participants are asked to identify the missing entry in a series of eight patterns, to complete a 3×3 matrix in which the items progressively change according to a particular (and often complex) characteristic. Participants had 20 min to work through the 36 items. The score was the number of items correctly answered. As demonstrated by Hamel and Schmittmann (2006), the score on the APM after 20 min is a reasonable predictor ($r = .74$) of the score when unlimited time is given.

In addition, we note that items in the APM are not all processed the same way (e.g., Carpenter, Just, & Shell, 1990; DeShon, Chan & Weissbein, 1995). DeShon et al. reported that a subset of items relies on visuospatial processes whereas another subset requires analytical processes. Because we were interested in whether spatial abilities tests draw on spatial mental imagery, we analysed separately the scores on the visuospatial subset of items and on the analytical subset, as defined by DeShon et al. We used these two scores to show that scores on our scanning task are related to performance on the spatial items of the APM, but not the analytic items.

OSIQ. The OSIQ (Blajenkova et al., 2006) is designed to assess individual differences in visual imagery preferences and experiences. Participants rate on a 5-point scale the degree to which they agree with each of 45 statements. The OSIQ has three different scales: an object imagery scale, a spatial imagery scale, and a verbal scale. A high score on the object scale indicates a preference for creating high-resolution and colourful visual mental images, whereas a high score on the spatial scale indicates a preference for using schematic mental images or images of relations between objects. The score on each scale is computed by adding the ratings of the participant on the 15 items of that scale. The questionnaire is untimed.

VVIQ. The VVIQ (Marks, 1973) is a self-report questionnaire, in which participants rate on a 5-point scale the vividness of their visual mental images. The questionnaire consists of 16 items. Participants had unlimited time to respond, and the score was the sum of the ratings.

Analysis of the image-scanning task

We began by conducting the same analyses as those in the two previous experiments. Defining outliers as in Experiments 1 and 2, 2.3% of the trials were considered as outliers. Then, we compared the slopes of the best fitting lines and ERs between the two levels of AoU to ensure that practice effects (due to performing the hardest condition last) did not affect the results. Finally, we conducted the correlational analysis between all the dependent variables in all the tasks and tests. Preliminary analyses did not reveal any effect of gender on the different dependent variables, and thus we pooled the data for males and females and do not address this factor in the following description of the results.

Yes trials. At each AoU level, RTs varied for different distances between the tip of the arrow and the nearest target dot, as revealed by 2 one-way repeated measures ANOVAs, $F(3, 141) = 45.72$, $MSE = 43{,}035.27$, $p < .0001$, $\eta_p^2 = .49$, for Level 1, and $F(3, 141) = 22.79$, $MSE = 40{,}034.46$, $p < .0001$, $\eta_p^2 = .33$, for Level 2. The best fitting functions computed by the method of the least squares revealed that RTs increased linearly with distance, with $F(1, 47) = 96.66$, $MSE = 61{,}231.53$, $p < .0001$, $\eta_p^2 = .66$, for Level 1, and $F(1, 47) = 138.74$, $MSE = 20{,}162.13$, $p < .0001$, $\eta_p^2 = .75$, for Level 2. In addition, scanning times and distance were highly correlated at both levels, $rs = .97$, $p < .05$, in both cases. As in Experiment 1 and Experiment 2, participants mentally scanned the distance between the tip of the arrow and the location of the dots memorized during the learning phase.

Comparisons of performance at the two AoU levels. If the AoU radii affect the degree of difficulty of the scanning task, then participants should make fewer errors and respond faster for Level 2 (easy condition) than for Level 1 (hard condition). And in fact, participants' scanning rate (as revealed by the steepness of the slopes of the best fitting lines) was significantly faster for Level 2 ($M = 70$ ms/cm) than for Level 1 ($M = 103$ ms/cm), $t(47) = -3.71$, $p < .0005$, $d = -0.56$. Participants were also faster for the No trials of Level 2 ($M = 1{,}562$ ms) than for those of Level 1 ($M = 1{,}750$ ms), $t(47) = -3.36$, $p < .005$, $d = -0.36$. Finally, participants made fewer errors for Level 2 ($M = 27.6\%$) than for Level 1 ($M = 31.5\%$), $t(47) = -3.36$, $p < .005$, $d = -0.41$. However, as reported in the two previous experiments, ERs were significantly different between the two AoU levels in the No trials (25% vs. 29.5%), $t(47) = -3.33$, $p < .005$, $d = -0.38$, but not on the Yes trials (32.4% vs. 30.2%), $t(47) = -1.38$, $p = .09$. The pattern of results replicated those reported in Experiment 2. In addition, we note that participants committed more errors and were slower in the condition in which the AoU radius was the smallest, which rules out a speed/accuracy trade-off.

Correlational analysis

In order to consider whether the ability to generate and to process precise spatial mental images is related to performance on spatial abilities tests, we examined the correlations among all dependent variables. If the measures in the image-scanning task reflect spatial mental imagery, and spatial mental images play a role in spatial cognition, then we expect correlations between measures of performance of the image-scanning task and the scores for the paper-and-pencil spatial tests. In addition, if spatial tests rely selectively on spatial imagery, and performance in the image-scanning task assesses specifically spatial mental imagery, then we do not expect a correlation between scores on the spatial tests or the measures of performance of the image-scanning task and the scores from the two self-report questionnaires that assess the quality of mental images of single objects (i.e., the VVIQ and the object scale of the OSIQ).

The correlational analysis included the slopes of the best fitting lines, the RTs (No trials) and accuracy (No trials) for each level of the image-scanning task, as well as the same measures when the two levels were considered together. We restricted our correlational analysis to the dependent variables that were affected by the AoU, and thus we excluded the ERs on the Yes trials in this analysis. In addition, we included the overall scores in the two spatial tests, the scores on the VVIQ, the scores on the object and spatial scales of the OSIQ, and the visual and analytic scores on the APM. A summary of the descriptive statistics is presented in Table 3 as well as the reliability coefficient for each of the dependent variables. For all the measures of the image-scanning task, we computed split-half reliability coefficients (odd–even trials). For the two spatial tests, the APM, the OSIQ, and the VVIQ, we reported the test–retest reliability coefficients estimated respectively by Ekstrom et al. (1976), Raven et al. (1998), and

Table 3. *Experiment 3: Summary of descriptive statistics and reliability coefficients for the dependent variables used in the correlational analysis*

Measure		*M*	*SD*	*Observed range*		*Reliability estimate*
				Minimum	*Maximum*	
IS task	Slope	69	41	11	126	.70
	Level 1	103	74	10	338	.55
	Level 2	70	40	5	151	.69
	Acc	72.2	11.1	51.2	92.3	.82
	Level 1	70.5	12.5	51.2	90.7	.74
	Level 2	75	12.4	54	97.5	.79
	RT	1,660	552	824	2,848	.88
	Level 1	1,750	504	795	2,989	.81
	Level 2	1,562	534	755	2,749	.85
PFB		19.42	10.5	2	37	.81
PF		12.9	4.3	5	20	.84
VVIQ		61.4	10.9	41	78	.87
OSIQ	Spatial	42.3	11.2	22	73	.78
	Object	45.7	10.5	20	68	.80
Raven	Total	21.4	5.6	5	31	.91
	Visual	8.4	2.9	2	13	
	Analytic	6.1	2.3	1	11	

Note: IS task: image-scanning task. Acc: accuracy on the No trials. RT: response time on the No trials. PFB = Paper Form Board. PF = Paper Folding. VVIQ = Vividness of Visual Imagery Questionnaire. OSIQ = Object Spatial Imagery Questionnaire.

Blajenkova et al. (2006). Finally, for the VVIQ, we reported Cronbach's alpha (see McKlevie, 1995). Overall, reliabilities of the measures were generally satisfactory, ranging from .74 to .94, $p < .01$ in all cases. However, the coefficients of reliability of the slopes of the best fitting lines were not as high as expected if these measures were reliable; coefficients of reliability ranged from .55 to .70, $p < .01$. Given the number of errors on the Yes trials, the RTs were probably more sensitive to outliers, which might have affected the reliability of the slopes.

Next, we examined the correlations between the dependent variables in the image-scanning task and the scores on the tests and questionnaires. Accuracy on the No trials of the image-scanning task was the only variable that correlated with the scores on the spatial tests: respectively, $r(46) = .34$, $p < .05$ with the PFB test, $r(46) = .39$, $p < .01$ with the PF test, and $r(46) = .47$, $p < .01$ with the visuospatial items of the APM (see Table 4). Interestingly, accuracy for the hardest level of AoU (Level 1), which provides the most sensitive measure of performance, was correlated with all spatial abilities measures, respectively, $r(46) = .36$, $p < .05$, with scores on the PFB test; $r(46) = .40$, $p < .01$, with scores on the PF test and $r(46) = .46$, $p < .01$ with scores on the visuospatial items of the APM. Taken together this pattern of correlations suggests that spatial tests rely at least partially on spatial mental imagery.

In addition, the scores on the spatial scale of the OSIQ were correlated with the overall accuracy on the No trials, $r(46) = .31$, $p < .05$, accuracy on the hardest level of the AoU (Level 1), $r(46) = .28$, $p < .05$, and scores on the PF and PFB tests: respectively, $r(46) = .34$, and $r(46) = .36$, ps $< .05$. Thus the cognitive style of the participants was related not only to their spatial abilities but also to their spatial mental imagery abilities.

Regarding whether spatial tests selectively involve spatial mental imagery, scores on the spatial tests did not correlate with scores on object-based mental imagery questionnaires (OSIQ object

Table 4. *Experiment 3: Matrix of correlations*

		Image-scanning task															
		Slope			*RT*			*Acc*			*Spatial tests*		*OSIQ*			*APM*	
		L1	*L2*	*All*	*L1*	*L2*	*All*	*L1*	*L2*	*All*	*PFB*	*PF*	*Spatial*	*Object*	*VVIQ*	*Visual*	*Analytic*
Image-scanning task																	
Slope	L1	1.00	.53**	.74**	.24	.33*	.30*	.18	.10	.16	−.19	−.09	−.25	−.02	−.07	−.01	−.06
	L2		1.00	.73**	.25	.34*	.35*	.28	.22	.28	−.06	.02	−.03	−.11	−.05	.19	.13
	All			1.00	.48**	.50**	.55**	.36*	.25	.35*	.00	.01	−.07	−.20	−.10	.13	.08
RT	L1				1.00	.72**	.90**	.22	−.08	.08	−.06	−.06	−.07	−.11	.16	−.04	.13
	L2					1.00	.90**	.33*	−.04	.16	.09	.13	−.17	−.11	.15	.15	.09
	All						1.00	.34*	−.02	.18	.08	.09	−.12	−.16	.11	.04	.21
Acc	L1							1.00	.57**	.89**	.36*	.40**	.28*	−.21	.08	.46**	.05
	L2								1.00	.89**	.25	.29*	.26	−.08	.00	.39**	.01
	All									1.00	.34*	.39**	.31*	−.16	.05	.47**	.04
Spatial tests	PFB										1.00	.83**	.34*	.03	.08	.42**	.23
	PF											1.00	.36*	.06	.14	.52**	.24
OSIQ	Spatial												1.00	−.28	.34*	.21	.17
	Object													1.00	.41**	.05	.03
	VVIQ														1.00	.13	.20
APM	Visual															1.00	.21
	Analytic																1.00

Note: L1: area of uncertainty (AoU) Level 1. L2: AoU Level 2. RT: response times. Acc: accuracy. PFB: Paper Form Board. PF: Paper Folding. OSIQ: Object Spatial Imagery Questionnaire. VVIQ: Visual Vividness Imagery Questionnaire. APM: Raven Advanced Progressive Matrices.
$^{*}p < .05$. $^{**}p < .01$.

and VVIQ, *r*s ranging from .01 to .14, *p*s $>$ .24 in all cases). In addition, accuracy on the No trials of the image-scanning task revealed individual differences on spatial mental imagery ability per se, given that this measure did not correlate with scores on the object scale of the OSIQ (*r*s ranging from −.16 to −.21, *p*s $>$ .15 in all cases), nor did it correlate with scores on the VVIQ (*r*s ranging from −.01 to .08, *p*s $>$ .50 in all cases), nor with scores on the analytic items of the APM (*r*s ranging from .01 to .05, *p*s $>$ .50 in all cases).

Discussion

In Experiment 3, we replicated the results of Experiment 2. When the radius of the AoU decreased, participants scanned at a slower rate and committed more errors in the No trials. Thus, performing the hardest condition last did not eliminate the practice effect of AoU. At an individual level, the participants' spatial abilities, as measured by objective spatial tests (Paper Folding, Paper Form Board, and scores on the visuospatial items of the Advanced Progressive Matrices), were related to the accuracy with which participants decided whether an arrow did not point at any of the dots. Performance on the spatial tests and the accuracy on the image-scanning task were also related to the cognitive style of the participants. Participants with stronger spatial abilities and participants who created the most accurate spatial representations of the patterns of dots were the ones who claimed to use spatial imagery in their everyday life (as revealed by their scores on the spatial scale of the Object Spatial Imagery Questionnaire). In addition, these correlations cannot be accounted for by participants' general intelligence nor by their object mental imagery abilities.

These results are of great interest because they provide the first evidence that spatial mental images are required to perform certain types of spatial tests. These findings support the role of spatial mental imagery in spatial cognition. In addition, the lack of relationship between participants' ability to create and process precise spatial mental images and their ability to generate and process object mental images supports the claims that (a) visual mental imagery is a collection of different abilities, and (b) object mental images and spatial mental images are created and processed by different neural systems.

However, counter to our hypothesis, the image-scanning speed of the participants, while affected by the radii of the AoU, was not related to their spatial abilities. Part of the explanation may be that the slopes of the best fitting lines for individual participants reflect not only the ability to generate a precise spatial mental image but also the ability to scan an image per se. Thus, although the type of spatial abilities measured in the spatial tests rely on spatial images, they have little to do with the ability to shift attention across (i.e., scan) a pattern in a mental image. In addition, the slopes of the best fitting lines exhibited greater measurement error than the accuracy measure on the No trials, as revealed by the moderate coefficients of reliability of the slopes. In fact, the slopes of the best fitting lines are not stable without a large number of experimental trials (see Borst & Kosslyn, 2008). We designed the task to be challenging, and hence participants made a large number of errors on the Yes trials—which increased the vulnerability of the RTs to outliers (which in turns affects the reliability of the slopes).

GENERAL DISCUSSION

In Experiments 1 and 2, at each of the AoU levels, as the distance between the arrow and the dot increased participants took more time to respond, which suggests that participants were scanning a spatial mental image that incorporates the metric properties of the array of dots. Such a pattern of results is usually interpreted as evidence that depictive representations were processed (for further discussion see Denis & Kosslyn, 1999). Because points in a spatial mental image are represented in a coordinate space, the logic of our paradigm was that it is possible to estimate the spatial error of the representation created by participants when they generated the image.

In Experiments 1 and 2, individual differences were observed both on the slopes and on the ERs, which suggests that we could use the image-scanning task as an objective way to measure individual differences in the precision of spatial mental imagery. In order to assess the quality of the spatial representation, we observed the effect of varying the AoU on performance. The logic underlying the variation of the AoU was that participants who generated the spatial mental image with the least precise locations of the dots should have taken longer and made more errors as the AoU decreased. In general, we found increasing numbers of errors and slower scanning when the radius of the AoU became smaller. Experiments 1 and 2 demonstrated that the image-scanning paradigm could measure effectively the efficiency with which one could generate precise spatial mental images. This finding was a prerequisite for investigating whether spatial abilities—as measured by spatial tests—rely on spatial mental imagery ability.

In Experiment 3, the pattern of correlations revealed that although performance on spatial tests (PF, PFB, and spatial items of the APM) was correlated with accuracy on the No trials of the image-scanning task, it was not related to scores on the VVIQ, ratings on the object scale of the OSIQ, or scores for the analytic items of the APM. Taken together, these results suggest that participants rely selectively on spatial mental images to perform at least certain types of spatial tasks. In addition, given that performance on the image-scanning task was not related to performance on the object-based task, we have evidence that the ability to generate a precise spatial mental image is different from the ability to generate mental images of single objects. This result is consistent with the finding that mental imagery is not unitary system but is a collection of abilities (see Kosslyn, 1994).

In short, the present studies have provided evidence that spatial mental imagery underlies at least some aspects of spatial ability. In addition, these findings suggest that the image-scanning paradigm may offer an objective method for studying individual differences in spatial mental imagery. To make progress in studying the role of individual differences in imagery in a host of tasks, ranging from learning to problem solving, we need ways to assess the various types of imagery. Only after we can assess imagery for shape and for colour, in addition to spatial location, are we likely to be able to predict which sorts of imagery would be most effective for a given person in a given task.

Original manuscript received 18 August 2009
Accepted revision received 09 February 2010
First published online 1 June 2010

REFERENCES

Bahill, A. T., & Stark, L. (1979). The trajectories of saccadic eye movements. *Scientific American*, *240,* 108–117.

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology*, *20,* 239–263.

Borst, G., & Kosslyn, S. M. (2008). Visual mental imagery and perception: Structural equivalence revealed by scanning processes. *Memory and Cognition*, *36,* 849–862.

Borst, G., Kosslyn, S. M., & Denis, M. (2006). Different cognitive processes in two image-scanning paradigms. *Memory and Cognition*, *34,* 475–490.

Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, *9,* 529–533.

Carpenter, P. A., & Just, M. A. (1986). Spatial ability: An information processing approach to psychometrics. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3). Hillsdale, NJ: Lawrence Erlbaum Associates.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97,* 404–431.

Danaher, B. G., & Thoresen, C. E. (1972). Imagery assessment by self-report and behavioural measures. *Behaviour Research and Therapy*, *10,* 131–138.

Dean, G. M., & Morris, P. E. (2003). The relationship between self-reports of imagery and spatial ability. *British Journal of Psychology*, *94,* 245–273.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. New York: Oxford University Press.

Denis, M., & Cocude, 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.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence*, *21,* 135–155.

Di Vesta, F. J., Ingersoll, G., & Sunshine, P. (1971). A factor analysis of imagery tests. *Journal of Verbal Learning and Verbal Behavior*, *10,* 471–479.

Dror, I., & Kosslyn, S. M. (1994). Mental imagery and aging. *Psychology and Aging*, *9,* 90–102.

Durndell, A. J., & Wetherick, N. E. (1976a). The relation of reported imagery to cognitive performance. *British Journal of Psychology*, *67,* 501–506.

Durndell, A. J., & Wetherick, N. E. (1976b). Reported imagery and two spatial tests. *Perceptual and Motor Skills*, *43,* 1050.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Ernest, C. H. (1977). Imagery ability and cognition: A critical review. *Journal of Mental Imagery*, *2,* 181–216.

Finke, R. A., & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *8,* 142–147.

Finke, R. A., & Pinker, S. (1983). Directional scanning of remembered visual patterns. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9,* 398–410.

Fuchs, A. F. (1976). The neuropsychology of saccades. In R. A. Monty & J. W. Senders (Eds.), *Eye movements and psychological processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.

Hamel, R., & Schmittmann, V. D. (2006). The 20-minute version as a predictor of the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement*, *66,* 1039–1046.

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception and Psychophysics*, *14,* 90–94.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: The MIT Press.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4,* 47–60.

Kosslyn, S. M., Brunn, J. L., Cave, K. R., & Wallach, R. W. (1984). Individual differences in mental imagery abilities: A computational analysis. *Cognition*, *18,* 195–243.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.

Kosslyn, S. M., Thompson, W. L., Gitelman, D. R., & Alpert, N. M. (1998). Neural systems that encode categorical vs. coordinate spatial relations: PET investigations. *Psychobiology*, *26,* 333–347.

Kozhevnikov, M., Kosslyn, S. M., & Shephard, J. M. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory and Cognition*, *33,* 710–726.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's g? In I. Dennis & P. Tapsfield (Eds.), *Human abilities*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Levine, D. N., Warach, J., & Farah, M. J. (1985). Two visual systems in mental imagery: Dissociation of "what" and "where" in imagery disorders due to bilateral posterior cerebral lesions. *Neurology*, *35,* 1010–1018.

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lorenz, C., & Neisser, U. (1985). Factors of imagery and event recall. *Memory and Cognition*, *13,* 494–500.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64,* 17–24.

Marks, D. F. (1977). Imagery and consciousness: A theoretical review from an individual differences perspective. *Journal of Mental Imagery*, *1,* 275–290.

McKlevie, S. J. (1995). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery*, *19,* 1–106.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, *6,* 414–417.

Mumaw, R. J., Pellegrino, J. W., Kail, R. J., & Carter, P. (1984). Different slopes for different folks: Process analysis of spatial aptitude. *Memory and Cognition*, *12,* 515–521.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt. Rinehart & Winston.

Pellegrino, J. W., & Kail, R. J. (1982). Process analyses of spatial aptitude. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). London: Lawrence Erlbaum Associates.

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 and Cognition*, *12,* 207–218.

Poltrock, S. E., & Agnoli, F. (1986). Are spatial visualization ability and visual imagery ability equivalent? In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3). Hillsdale, NJ: Lawrence Erlbaum Associates.

Poltrock, S. E., & Brown, P. (1984). Individual differences in visual imagery and spatial ability. *Intelligence*, *8,* 93–138.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. I. Pick, Jr. & E. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 137–157). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25,* 157–238.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual Section 4: Advanced Progressive Matrices*. Oxford, UK: Oxford Psychologists Press.

Rehm, L. P. (1973). Relationships among measures of visual imagery. *Behavioural Research and Therapy*, *11,* 265–270.

Richardson, A. (1977). The meaning and measurement of memory imagery. *British Journal of Psychology*, *68,* 29–43.

Segal, S. J., & Fusella, V. (1969). Effect of imagining on signal-to-noise ratio with varying signal conditions. *British Journal of Psychology*, *60,* 459–464.

Sheehan, P. W., & Neisser, U. (1969). Some variables affecting the vividness of imagery recall. *British Journal of Psychology*, *60,* 71–80.

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.