

# Functional Equivalence of Spatial Representations Derived From Vision and Language: Evidence From Allocentric Judgments

Marios N. Avraamides and Jack M. Loomis  
University of California at Santa Barbara

Roberta L. Klatzky  
Carnegie Mellon University

Reginald G. Golledge  
University of California at Santa Barbara

Past research (e.g., J. M. Loomis, Y. Lipka, R. L. Klatzky, & R. G. Golledge, 2002) has indicated that spatial representations derived from spatial language can function equivalently to those derived from perception. The authors tested functional equivalence for reporting spatial relations that were not explicitly stated during learning. Participants learned a spatial layout by visual perception or spatial language and then made allocentric direction and distance judgments. Experiments 1 and 2 indicated allocentric relations could be accurately reported in all modalities, but visually perceived layouts, tested with or without vision, produced faster and less variable directional responses than language. In Experiment 3, when participants were forced to create a spatial image during learning (by spatially updating during a backward translation), functional equivalence of spatial language and visual perception was demonstrated by patterns of latency, systematic error, and variability.

People typically learn about space by means of direct perception. By viewing, hearing, touching, or moving around objects in their environment, they can form spatial representations of their physical surroundings. However, mental representations of space are not formed exclusively through direct perceptual input. They can also be constructed indirectly by means of symbolic media such as maps (e.g., Richardson, Montello, & Hegarty, 1999), diagrams (e.g., Bryant, Lanca, & Tversky, 1995; Bryant & Tversky, 1999), and language (e.g., Avraamides, 2003; Denis & Zimmer, 1992; Ferguson & Hegarty, 1994; Taylor & Tversky, 1992).

Spatial representations derived from indirect sources have been shown to have properties similar to those formed through perception. In the case of language, the vast literature of situation models (van Dijk & Kintsch, 1983) or mental models (Johnson-Laird, 1983, 1996) have documented that the properties of physical environments are preserved in mental representations that are formed through language (see Zwaan & Radvansky, 1998, for a review). For example, mental representations derived from language maintain the metric properties of real environments (Glenberg, Meyer, & Lindem, 1987; Morrow, Greenspan, & Bower,

1987). This is supported by evidence that has shown that the latency to complete mental scanning between two objects in described scenes varies as a function of the interobject distance (Kosslyn, Reiser, Farah, & Fliegel, 1983). This finding was also obtained in a study by Denis and Cocude (1989) in which participants mentally scanned the distances among landmarks of a verbally described island. In addition, Denis and Cocude (1989) demonstrated that this dependence of mental scanning latency on interobject distance was also obtained when participants studied maps of the island instead of reading verbal descriptions. Finally, Mellet, Bricogne, Crivello, Mazoyer, Denis, et al. (2002) showed, using the positron emission tomography methodology, that mental scanning of map that was built from text engaged the same parieto-frontal network that is active when scanning representations built from vision (Mellet, Bricogne, Tzourio-Mazoyer, Ghaem, Petit, et al., 2000). However, it should be noted that both of these studies involved participants with high visuo-spatial imagery, who would be more likely to convert text into analog spatial representations.

The similar results with maps and linguistic descriptions in the study by Denis and Cocude (1989), in conjunction with the evidence for common neural correlates, suggest the possibility that although spatial representations can be constructed through various means, at some point the source of encoding becomes unimportant. That is, once they are well formed, the representations of spatial layouts derived from vision, audition, touch, language, pictures, maps, diagrams, and so on are all functionally equivalent and perhaps even identical (i.e., amodal; Bryant, 1997; De Vega, Cocude, Denis, Rodrigo, & Zimmer, 2001; Loomis et al., 2002). This hypothesis is further reinforced by the similarity of findings reported by studies that required localization of objects in scenes that were learned perceptually (Hintzman, O'Dell, & Arndt, 1981)

---

Marios N. Avraamides and Jack M. Loomis, Department of Psychology, University of California at Santa Barbara; Roberta L. Klatzky, Department of Psychology, Carnegie Mellon University; Reginald G. Golledge, Department of Geography, University of California at Santa Barbara.

This research was supported by National Eye Institute Grant EY09740. We thank Jared Shields for valuable assistance with data collection and analysis. We also thank Michael Masson, Michel Denis, and Timothy McNamara for providing useful comments on a draft of this article.

Correspondence concerning this article should be addressed to Jack M. Loomis, Department of Psychology, University of California, Santa Barbara, CA 93106-9660. E-mail: loomis@psych.ucsb.edu

or through texts (e.g., Avraamides, 2003; De Vega & Rodrigo, 2001; Franklin & Tversky, 1990).

Despite the evidence for functional equivalence among spatial representations constructed from various sources, a number of studies have provided data suggesting that differences between spatial representations derived from direct perception and language might exist. For example, using the mental scanning paradigm with described targets, Denis and Cocude (1997) were unable to find any spatial biases when scanning toward salient as compared with toward neutral landmarks. Asymmetries in spatial responses directed toward salient and nonsalient locations are a typical finding with real-world environments (McNamara & Diwadkar, 1997; Newcombe, Huttenlocher, Sandberg, Lie, & Johnson, 1999). Moreover, spatial representations derived from language do not seem to be as vivid as ones derived from perception. For example, Federico and Franklin (1997) found that information about spatial relations was retained longer in memory when it was encoded from pictures rather than text. Finally, the studies by Mellet et al. (2000, 2002) revealed that despite the common activation of the parieto-frontal network, language and vision activated (during mental scanning) other areas that were specific to the input; language selectively activated the angular gyrus and Broca's and Wernicke's areas (Mellet et al., 2002), whereas vision selectively activated the medial temporal lobe (Mellet et al., 2002). This result suggests that a trace of the input modality was still present well after encoding was completed.

It also should be pointed out that the arguments for functional equivalence of spatial representations do not posit that these representations are formed with the same ease across modalities. In fact, Klatzky, Lippa, Loomis, and Golledge (2002) showed that participants took longer to form a stable spatial representation when the targets were learned linguistically than when learned through direct perception. In this study, Klatzky et al. measured how many trials participants would need to learn the azimuths of five targets through vision, spatial audition (three-dimensional [3-D] sound), and spatial language (e.g., "1 o'clock, 6 feet"). Results revealed a disadvantage for spatial language with five-target trials even when proprioceptive cues, optic flow information, and differential binaural signals were controlled for (Experiment 2).

Despite the disadvantage of spatial language documented by Klatzky et al. (2002), a study by Loomis et al. (2002) provided evidence that once spatial representations are formed, they appear to be similar regardless of the input modality. In that study, participants first encoded the location of a target either through 3-D sound or spatial language and then walked to the target without vision along direct and indirect paths. Results showed that, for both input modalities, stopping points for direct and indirect paths converged remarkably. This result suggests that, with both modalities, participants converted the information about the target into a spatial image, which they updated continuously during their movement. Loomis et al. (2002) concluded that the spatial representations formed from 3-D sound and spatial language are functionally equivalent or very nearly so. A subsequent study (Klatzky, Lippa, Loomis, & Golledge, 2003) that involved the updating of multiple targets provided some additional support for the functional equivalence hypothesis, although there were small reliable differences between spatial language and the two perceptual modalities (vision and spatial hearing) in terms of updating.

On the basis of these results, Loomis et al. (2002) proposed a two-stage model of stimulus encoding and spatial updating. According to this model, an *encoding stage*, which can receive input from any modality, processes the stimulus and creates a spatial image. Then, an *updating stage* is responsible for keeping the spatial image up-to-date; that is, it computes new egocentric relations whenever egocentric relations change (as in the case of observer movements). Because it assumes that spatial updating takes place independently of the source of the spatial image, the model accounts for the similarities in updating across modalities that are reported by the empirical studies. The model posits that the spatial images formed from various inputs are functionally equivalent, but as Klatzky et al. (2003) pointed out, it does not specify what the nature of the spatial image is. One possibility is that different inputs converge to an image of visual format which is updated on the basis of imagined optic flow (Rieser, 1989). However, a finding from Loomis et al. (2001) that congenitally blind participants can update equivalently with 3-D sound and spatial language casts doubt on this hypothesis. An alternative hypothesis is that the spatial image is amodal. This hypothesis was proposed by Bryant (1997), who argued that a common spatial representation system, which receives input from perception or language, constructs amodal spatial images that represent information in a format that is neither perceptual nor linguistic.

The previous studies that directly compared spatial images from perception and spatial language (e.g., Klatzky et al., 2003; Loomis et al., 2002) have tested only egocentric relations (with or without spatial updating). However, these spatial images should, in principle, also convey information that is not egocentric. Specifically, a spatial image that codes the locations of targets relative to the observer (i.e., is *egocentric*) should allow the computation of object-to-object relations that are *allocentric*, that is, that refer one object to a coordinate system defined by another object (Klatzky, 1998; see also Gallistel, 1990). Allocentric relations comprise both distance and directional information: *allocentric interpoint distances* refer to the distances between pairs of objects, and *allocentric interpoint bearings* to the angles formed by a line from one object to another, relative to an axis of reference (Klatzky, 1998). In the present study, we attempt a more rigorous examination of the functional equivalence hypothesis by comparing people's ability to report allocentric relations after they learned spatial layouts from vision and spatial language. The participant's task involved a learning phase, in which the spatial layout of four objects was presented until the subject recalled it with criterion accuracy, and an allocentric report phase, in which the relations among the objects were reported. The learning phase duration was measured in trials, and the allocentric report phase was measured by response latency; errors were also recorded.

Figure 1 shows a model that specifies three phases involved in reporting allocentric relations of objects that were learned from a single vantage point. The phases are (a) encoding the presented

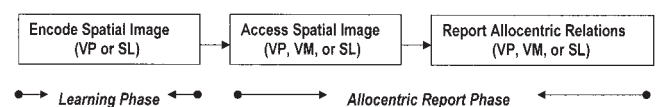


Figure 1. Model of reporting allocentric relations. VP = visual perception; VM = visual memory; SL = spatial language.

spatial information into a spatial image, (b) accessing the spatial image when queried, and (c) reporting the allocentric relations. Figure 1 shows three circumstances under which these processes occur in our experiments. Encoding may occur by visual perception (VP) or by means of spatial language (SL). In the latter condition, subjects are verbally told the locations of individual objects that they do not see. A third condition used here is visual memory (VM) in which the spatial image is created by visual perception, but subsequent reporting is done without vision.

We assume that initially, an egocentric spatial image is formed, from which allocentric relations are computed (see Burgess, Jeffery, & O'Keefe, 1999). It is possible that the allocentric relations are determined immediately, but another possibility is that the computation is not performed until the allocentric report is called for.

Our initial hypothesis follows the model of Loomis et al. (2002) in hypothesizing functional equivalence of spatial images formed through spatial language and perception once they are encoded into memory. The model, together with additional assumptions based on empirical findings, leads to a set of initial predictions, as follows.

#### Prediction 1

With respect to encoding the spatial image, VP will show faster learning than SL. We have confirmed this prediction in previous studies (Klatzky et al., 2002, 2003). Along with faster learning, VP may show less error at the end of learning. The relation of VM encoding to VP and SL is less clear, but it is likely that because of memory loss, VM will show some decrement in learning rate and precision relative to VP but possibly not as much as SL.

#### Prediction 2

This prediction concerns the duration of the allocentric report phase, which comprises two components. Accessing the spatial image is the first component. In this respect, VP should be faster and more accurate than VM or SL, because VP offers direct perception, whereas VM and SL require retrieving the image in memory. This assumption is supported by evidence from a number of studies in the motor literature that show that movement accuracy toward targets is reduced when the response is executed after a period in which targets are nonvisible (Elliott & Madalena, 1987; Westwood, Heath, & Roy, 2003). Furthermore, Westwood et al. (2003) showed that the variability of the reaching endpoint increased linearly with the delay of response after vision was occluded.

A question arises, however, as to whether VM and SL will be equivalent in the access phase. The model specifies equivalence if the two conditions both involve accessing a spatial image that has previously been encoded into memory. However, it is possible that people exposed to verbal descriptions of an object's location will defer the encoding into a spatial image until it is required by some task, such as walking to the target (Loomis et al., 2002) or reporting its location from a new vantage point (Klatzky et al., 2003). In that case, SL should be slower than VM in the access stage, because the access phase for SL includes deferred spatial encoding.

A second component of the allocentric report phase is computing the allocentric relations from the accessed image if they are not

already available. If the computation is necessary at this phase, VP will again be faster than VM or SL because of the perceptual basis for processing. This will merely increase the allocentric-report advantage for VP that was predicted on the basis of accessibility. More importantly, VM and SL will be equivalent in the time for this computation, and hence differential computation time will not underlie any differences in allocentric report latency. It is this equivalence of function for VM and SL, once spatial images have been encoded into memory and accessed for processing, that lies at the heart of the model's assumption of functional equivalence.

Consideration of both components together—accessing the spatial image and computing allocentric relations—leads to Prediction 2 about the response latency in the allocentric report phase (which was measured here only for the pointing component of the report). Specifically, VP will be faster than VM or SL, which will be equivalent. However, this prediction will fail, and SL will be slower than VM, if the formation of a spatial image from language is deferred.

#### Prediction 3

Functional equivalence of two modalities implies that they will be affected in the same way by variations in computing allocentric relations, which arise from variations in target pairs within the spatial image. Stimulus-based variation should affect both systematic bias and duration of processing. This leads to the prediction that VM and SL will show correlations with respect to both signed error and latency where the unit of observation is a target pair. Correlations should be reduced, however, if SL showed higher pointing latency than VM, indicating deferred image formation.

#### Prediction 4

We assume that each processing stage has noise associated with it, and the total noise accumulates with the number of processing stages. Indeed, if the noise in each stage is independent of the noise in any other stage, the total noise ought to be the sum of the constituent noises. This total noise is manifest as the variability (variance) of the responses. Ideally, this variability would be obtained for each participant, but lacking this, we instead used between-participants variability, computed for each stimulus pair. Instead of reporting variances, however, we chose to report standard deviations.<sup>1</sup>

By the assumption of cumulative noise, standard deviations of responses should increase if an additional stage is added, as would

<sup>1</sup> The standard deviation of signed error can be understood as a measure of noise by partitioning the error for any one target pair and participant into two additive components: (a) the systematic signed error represented by the mean ( $M$ ), and (b) the participant's deviation from that mean ( $\Delta$ ). If these errors are squared and summed across  $N$  participants, the sum is equal to

$$N \times M^2 + \sum \Delta^2,$$

(the cross product term in the participants' individual squared error drops out when they are summed, because the sum of deviations from the mean is zero). Thus the sum of squared error has two components, systematic and due to individual variations, or noise. The noise term  $\sum \Delta^2$  is proportional to  $SD^2$  (it is variance  $\times [N - 1]$ ). It is worth noting that although latency was consistently correlated with noise ( $SD$ ) across target pairs, it was not so correlated with the systematic component of signed error.

occur if the formation of an image from spatial language were deferred. This leads to a prediction with two contingencies: If there is a greater latency for SL relative to the other modality conditions, indicating such deferral, there should also be a greater standard deviation. If, on the other hand, there is no latency difference between VM and SL, indicating functional equivalence, then noise arising during the allocentric report stage should come from equivalent sources. This will lead to the two modalities having equal standard deviations that are correlated across target pairs.

### Prediction 5

More generally, there is the possibility that noise (standard deviation) is closely linked to response latency. If so, two modalities that are functionally equivalent, as judged by having equivalent response latencies, should also have equivalent standard deviations, and modalities that differ in latency should differ in standard deviation.

In short, mean latency, errors, and standard deviations, along with correlations between modalities and measures, provide a means of testing five predictions based on the hypothesis of functional equivalence. The dependent measures are most meaningful when they deviate from baseline values in which case intermodal correlations in the patterns across stimuli can be assessed along with intermodal differences in mean values. Strong support for the hypothesis obtains when there are systematic errors in these judgments and the pattern of systematic errors is the same for the two modalities. Other strong evidence for the hypothesis obtains when there are systematic variations in pointing latency as a function of an independent variable (e.g., interpoint distance for each pair of targets), and these systematic variations are the same for the two modalities.

In Experiment 1, we compared performance across SL, continuous VP, and VM. In the vision conditions, the objects were presented simultaneously. In Experiments 2 and 3, we compared spatial language with a visual–memory condition in which objects were presented sequentially. Because Experiments 1 and 2 indicated that SL was slower than VM, a result which could be due to deferred encoding of the spatial image in the SL condition, Experiment 3 added a further requirement that was intended to induce complete encoding of the spatial image during the learning phase. This allowed a full test of functional equivalence.

### Experiment 1

This experiment compared memory for multiple targets that were learned through VP and egocentric SL. After learning the locations of four targets participants were asked to report the allocentric direction and distance of pairs of targets. Allocentric direction responses were made with a pointing device, and allocentric distances were reported verbally. Two vision conditions were included. In the VM condition, participants learned the targets visually, but, as in SL, made all responses without vision. In the VP condition, targets were perceptually available at the time when participants executed their responses.

### Method

#### Participants

Twenty-four (14 male, 10 female) students of introductory psychology classes at the University of California, Santa Barbara, participated in the

experiment in exchange for course credit. Two of the participants did not understand the instructions on what to report and were therefore replaced.

#### Design

The experiment followed a within-subjects design, with every participant performing the task under three conditions: VP, VM, and SL. The order of conditions was counterbalanced across subjects.

#### Experimental Setup

A 7.0 m × 5.4 m room was used to conduct the experiment. There were three layouts of targets, which were determined by randomly sampling four egocentric directions and four distances from a pool containing angles deviating  $-90^\circ$ ,  $-60^\circ$ ,  $-30^\circ$ ,  $30^\circ$ ,  $60^\circ$ , or  $90^\circ$  from the subject's facing direction (negative angles indicating angles to the left of the facing direction) and egocentric distances of 0.91 m, 1.83 m, 2.73 m, 3.66 m, and 4.57 m. The polar coordinates of Layout A were:  $-90^\circ/3.66$  m,  $-60^\circ/1.83$  m,  $30^\circ/2.73$  m, and  $90^\circ/0.91$  m. The coordinates of Layout B were:  $-90^\circ/1.83$  m,  $-60^\circ/2.73$  m,  $30^\circ/4.57$  m and  $90^\circ/0.91$  m. Those of Layout C were  $-30^\circ/0.91$  m,  $30^\circ/4.57$  m,  $60^\circ/1.83$  m, and  $90^\circ/2.73$  m. These three layouts were used for all participants, but their assignment to the three conditions was counterbalanced across participants. Three sets of object labels—each containing the names of four familiar objects—were constructed and, for each participant, were randomly assigned to the three layouts. Object labels were printed on cards that were mounted on microphone stands at a height of 1.07 m. The pointer was a rod that rotated about the center of its long axis over a full circle. The participant responded by rotating the rod to match the bearing from one object to another. A button on the pointer was interfaced with a stopwatch that was used to record latencies for pointing responses.

#### Procedure

The experiment was divided into two phases: a learning phase in which participants learned the locations of the targets and a test phase in which they reported allocentric directions and distances.

*Learning phase.* Participants performed learning trials in which they were first exposed to the four targets of the layout and then pointed to each target and estimated its egocentric distance. Participants were given unlimited time to inspect the layouts in the VP and VM condition. In the SL condition, participants were blindfolded during the exposure period, and an experimenter described each target's position by specifying egocentric directions in clock positions and distances in feet (e.g., "cat, 3 o'clock, 9 feet"). The order in which objects were introduced was random for each participant, and repetitions of target positions were performed as requested by the participant. After the locations were encoded, an experimenter probed participants with the names of the targets in random order. After hearing the name of a target, participants rotated the pointer to the direction of the target and then provided an estimate of its distance. For VM and SL, participants were blindfolded during this test phase. Before each pointing response, an experimenter rotated the pointer into alignment with the participant's sagittal axis. Learning trials continued with alternating exposures and tests until completion of three trials in which the absolute pointing error, averaged across objects, was less than  $15^\circ$  and the reported egocentric distances achieved a rank correlation of .75 or higher.

*Test phase.* After the learning criterion was met, participants were probed with pairs of targets (e.g., "cat, baby") and they were asked to rotate the pointer into alignment with the allocentric direction of the two targets. As soon as the experimenter finished naming the second target, he started the time on the stopwatch. Participants pressed the button on the pointer to stop the recording of time as soon as they completed their pointing response. The pointer was aligned with the participant's frontal plane before each response. After each pointing response, participants provided

a verbal estimate of the distance between the two targets in feet. For the VM and SL conditions, participants wore blindfolds throughout the duration of this phase.

## Results

### Learning Phase

The measures obtained during the learning phase were trials to criterion and error at the end of learning. Only pointing error is considered, because given the verbal distance response, subjects in the SL condition could achieve highly accurate distance responses purely by repeating back the verbal information they had memorized. Prediction 1 was confirmed in that VP showed the fastest learning and greatest precision, whereas SL showed the slowest learning and least precision.

A repeated-measures analysis of variance (ANOVA), using the encoding modality as a within-subjects variable, was conducted on the number of trials needed to reach the learning criterion. The analysis revealed that, as predicted, participants needed the most trials to reach the criterion in the SL condition (4.42 trials;  $SD = 1.10$ ) and the fewest in the VP condition (3.00 trials,  $SD = 0$ , vs. 3.21 trials,  $SD = 0.42$ , for the VM condition),  $F(2, 46) = 30.40$ ,  $MSE = 0.46$ ,  $p < .001$ . All three pair-wise comparisons (using two-tailed  $t$  tests) were statistically reliable; all  $ps < .05$ . Most trials that failed to reach the learning criterion (4 out of 5 in VM and 21 out of 34 in SL) involved categorical errors; that is, participants responded to a target with the direction–distance of a different target.

Statistical analyses of absolute error were performed using participants' pointing responses from the third successful learning trial. Absolute direction error, measured as the absolute angular deviation of the pointing response from the correct response, was smallest in VP ( $3.04^\circ$ ,  $SD = 1.42$ ), intermediate in VM ( $5.08^\circ$ ,  $SD = 2.89$ ) and greatest in SL ( $7.77^\circ$ ,  $SD = 3.54$ ),  $F(2, 46) = 16.30$ ,  $MSE = 8.25$ ,  $p < .001$ ; all pairwise  $ps < .05$ .

Finally, participants' pointing responses were used to compute psychophysical functions to relate responses to correct values. A slope of 1.0 and an intercept of 0 would constitute fully accurate responses; departures indicated signed error. The slopes of the three functions were .999 ( $SD = .00$ ) for VP, .998 ( $SD = .00$ ) for VM, and .994 ( $SD = .00$ ) for SL,  $F(2, 46) = 8.16$ ,  $MSE = .00$ ,  $p < .05$ . Pairwise comparisons involving VP were significant,  $ps < .05$ . The contrast between VM and SL was only marginally significant,  $p = .07$ . The intercepts of the functions were  $-.25^\circ$  ( $SD = 3.16$ ) for VP,  $-.54^\circ$  ( $SD = 6.24$ ) for VM, and  $.45^\circ$  ( $SD = 7.03$ ) for SL and did not differ significantly from each other.

### Test Phase

For the test phase, several measures were relevant, including pointing response latency, signed error for pointing and distance responses, and the standard deviation of responses. Recall that the key predictions (and corresponding prediction numbers) from the introduction were (a) VP should show the fastest allocentric report, with VM equivalent to SL; (b) there should be correlations (over target pairs) between functionally equivalent modalities with respect to both signed error and pointing latency; (c) VM and SL should differ in standard deviation if they differ in latency, whereas the standard deviation in VM and SL should be equal and

correlated if the conditions do not differ in latency; and (d) VP and VM should differ in pointing standard deviation if they differ in latency.

**Pointing latency.** Overall, latencies for pointing responses were longer in the SL condition, intermediate in the VM condition, and shortest in the VP condition (Figure 2, top).<sup>2</sup> These differences were significant,  $F(2, 46) = 37.41$ ,  $MSE = 1.95$ ,  $p < .001$ ; all pairwise  $ps < .001$ . The difference between VM and SL was not as predicted and suggests that formation of the spatial image may have been deferred in the SL condition.

**Accuracy.** Figure 3 shows participants' average pointing responses as a function of the correct allocentric direction, thus indicating the signed error tendencies. (We do not report ANOVAs on signed error, because means are subject to canceling effects from target pairs with different bias directions.)

Figure 2 (bottom) shows pointing standard deviation, averaged over the target pairs. The noise measure was smallest for VP, intermediate for VM, and greatest for SL,  $F(2, 46) = 16.97$ ,  $MSE = 28.05$ ,  $p < .001$ ; all pairwise  $ps < .05$ . Thus, as predicted, latency differences were accompanied by matching differences in noise. Besides the correlated changes between latency and SD across modalities, we looked for within-modality correlations between the two measures.<sup>3</sup> Table 1 gives the correlation between standard deviation and latency for the pointing judgments over the 12 target pairs, for each modality in Experiment 1 (as well as in the experiments to follow). The average correlation for Experiment 1 was .63.

Figure 4 presents reported allocentric distance as a function of correct allocentric distance. Distance standard deviation was smaller for VP ( $.54$ ,  $SD = .31$ ) than for both VM ( $.77$ ;  $SD = 0.37$ ) and SL ( $.96$ ,  $SD = 0.32$ ),  $F(2, 46) = 8.21$ ,  $MSE = 0.13$ ,  $p < .01$ ; pairwise  $ps < .05$ . VM and SL were marginally different,  $p = .10$ .

**Correlations between modalities.** Table 2 shows the inter-modal correlational patterns across experiments for latency, signed error, and standard deviation. There was a significant latency

<sup>2</sup> For pointing latency, as well as standard deviation, we also analyzed the data separating participants who did not have SL as the first condition in the testing order. No changes were found in the pattern of results.

<sup>3</sup> We report here two types of correlations: (a) within a given modality, between two measures (e.g., correlating latency and noise within VM), and (b) for a given measure, between two modalities (e.g., correlating mean latency between VM and SL). Strictly speaking, our design violates the assumption that observations in a correlation be independent, because participants are counterbalanced across layouts. However, neither correlation appears to be compromised, as follows: (a) Considering correlations between latency and noise (within a modality), some participants will contribute to some target pairs, and other participants will contribute to the remaining pairs. To confirm that observed correlations were not due to interparticipant variation (e.g., slower participants were noisier), we measured the correlations over target pairs within single subgroups of participants and found the same patterns as are reported in Table 1. (b) Considering correlations between two modalities, for example, between mean latency for VM and SL, different subgroups will contribute to the two means. For example, Participant 1 might contribute to the VM mean and Participant 2 might contribute to the SL mean for one observation, whereas for another, Participant 1 might contribute to the SL mean and Participant 2 might contribute to the VM mean. Spurious correlations between modalities will not arise from participant differences, because the same participants do not contribute to the two means within a correlation.

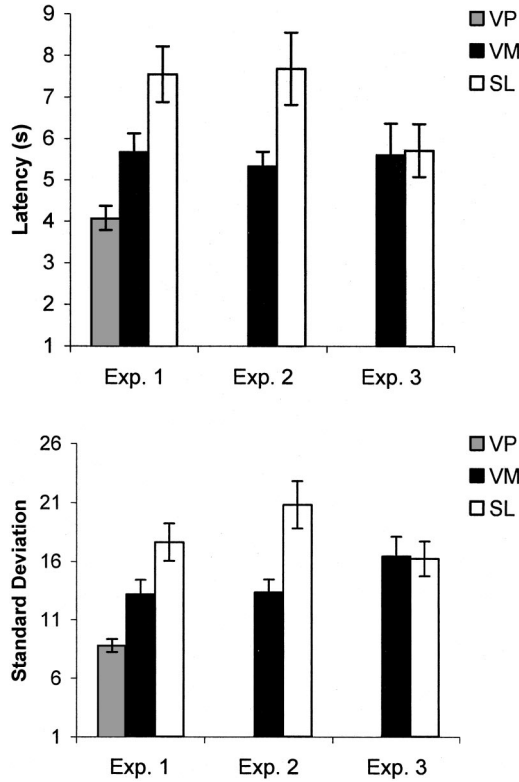


Figure 2. Average latency and standard deviation of signed pointing error in the test phase of each experiment as a function of input modality. Bars represent standard error of the mean. Exp. = Experiment; VP = visual perception; VM = visual memory; SL = spatial language.

correlation only between VP and VM but not, as would be predicted by functional equivalence, between VM and SL. The same was true of signed pointing error and standard deviation for pointing, although VM and SL correlated significantly for both distance

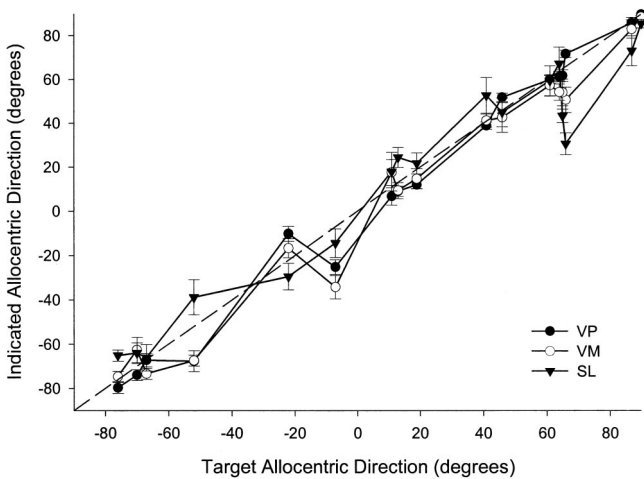


Figure 3. Reported allocentric direction response as a function of physical allocentric directions in Experiment 1. VP = visual perception; VM = visual memory; SL = spatial language.

Table 1  
Correlation Between Response Latency and Standard Deviation Across Target Pairs by Experiment

Experiment	Modality					
	VP		VM		SL	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Experiment 1	.43	.08	.82	.00	.64	.01
Experiment 2			.63	.03	.60	.04
Experiment 3			.60	.04	.40	.19

Note. The *df* was 15 for Experiment 1 and was 10 for Experiments 2 and 3. VP = visual perception; VM = visual memory; SL = spatial language.

signed error and distance standard deviation. Overall, this pattern suggests that VM and SL had substantially different sources of duration, systematic error, and noise.

Discussion

The main finding from this experiment is that performance was better when spatial layouts were encoded from vision—whether testing was with or without sight—than after encoding from spatial language. As expected, learning the layout was fastest and most accurate (particularly for pointing) in the VP condition, and it was slowest and produced greatest pointing error in the SL condition. In the allocentric-relations test for pointing, relative to VM and SL, the VP condition showed faster latency and lower noise than VM or SL. In turn, judging allocentric direction via VM was faster than SL and showed lower noise. The pattern of means for distance standard deviation followed that of pointing and latency.

The advantage of the VP condition overall was expected, given the model introduced in Figure 1. However, the advantage of VM over SL contrasts with studies that have shown very similar performance between the two modalities in spatial-updating tasks (Klatzky et al., 2003; Loomis et al., 2002). As the introduction

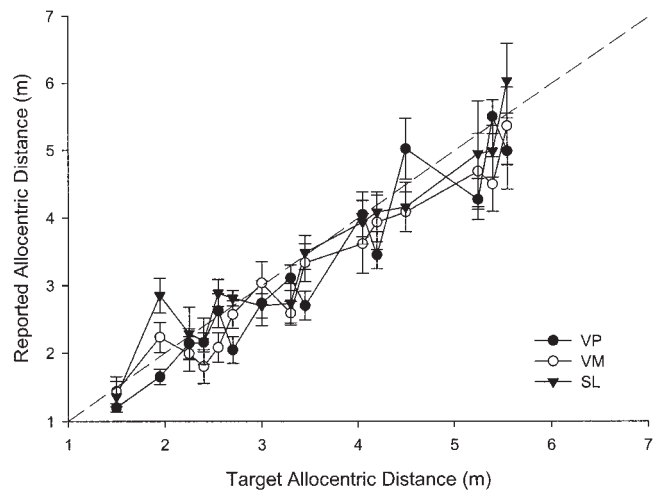


Figure 4. Reported allocentric distance as a function of physical allocentric distance in Experiment 1. VP = visual perception; VM = visual memory; SL = spatial language.

Table 2  
Correlations Between Modalities With Respect to Pointing Latency, Mean Error, and Standard Deviation of Pointing and Distance Responses

Experiment	Pointing response latency	Signed pointing error	SD of pointing responses	Signed distance error	SD of distance responses
Experiment 1					
VP-VM	.52*	.64**	ns	ns	ns
VM-SL	ns	ns	ns	.68**	.66**
Experiment 2					
VM-SL	.76**	ns	.76**	ns	ns
Experiment 3					
VM-SL	.71**	.93**	.57*	.67*	ns

Note. The *df* was 14 for Experiment 1 and was 10 for Experiments 2 and 3. VP = visual perception; VM = visual memory; SL = spatial language. \*  $p < .05$ . \*\*  $p < .01$ .

indicated, equivalence of SL to VM was not expected if participants who were exposed to spatial language failed to form a spatial image during the learning phase, instead waiting until they were queried about the allocentric relations in the test phase. This delay would increase the duration of the test phase for SL due to increased time to access the spatial image. It should also, by assumptions above, add noise, as measured by the between-subjects variability, as was observed. Moreover, the incorporation of image formation into the test phase could undermine correlations between VM and SL by introducing sources of systematic bias, noise, and latency beyond those attributable to computing the allocentric relations.

Another possibility, however, is that the VM condition had an advantage over SL because of the fact that the objects were presented simultaneously with vision but sequentially with language. The simultaneity of vision could create a more accessible spatial image and could promote the computation of allocentric relations during the learning phase, prior to the test. Either of these effects would shorten the latency for the VM condition during the test phase. To eliminate such advantages for visual memory, Experiment 2 introduced a VM condition that was sequential.

### Experiment 2

In Experiment 2 we presented the objects in the spatial layout one after another. If the advantage for the VM condition over SL is due to simultaneous availability of the objects during encoding, this advantage should be eliminated.

### Method

#### Participants

Sixteen (10 male, 6 female) students of introductory psychology classes at the University of California, Santa Barbara, participated in this experiment in exchange for course credit. One of the participants failed to understand the experimental instructions on what direction to indicate and his data were therefore discarded from all analyses.

#### Design and Procedure

Experiment 2 was identical to Experiment 1 with a few notable exceptions. First, only the VM and SL conditions were tested. Two layouts were

used (Layouts B and C from Experiment 1), and their assignment to modalities was counterbalanced across subjects. In contrast to Experiment 1, the presentation of targets in the learning phase was sequential for both SL and VM. Throughout the learning phase, participants wore a blindfold; in VM, they removed the blindfold only after the experimenter had placed one of the targets in the appropriate location. They were given unlimited time to inspect the target and then to put the blindfold back on. The experimenter removed the target (with its microphone stand) and placed a new one at a different location. This procedure was repeated until all four targets were viewed. The SL condition was as in Experiment 1.

### Results

#### Learning Phase

As in Experiment 1, participants needed more trials to reach the learning criterion with SL (4 trials;  $SD = 1.13$ ) than with VM (3.14 trials;  $SD = .35$ ),  $t(14) = -2.98$ ,  $p < .05$ . Again, most trials failing to reach the learning criterion (2 of 2 in VM and 10 of 15 in SL) involved categorical errors. However, in contrast to Experiment 1, absolute error for pointing responses did not differ significantly between the two modalities,  $t(14) = -0.50$ ,  $p = .63$ ;  $6.97^\circ$  ( $SD = 3.61$ ) and  $7.53^\circ$  ( $SD = 2.95$ ), respectively, for VM and SL.

Psychophysical functions relating pointing responses to correct values had statistically equal slopes for VM (.996,  $SD = 0.01$ ) and for SL (.992,  $SD = .02$ ),  $t(14) = 1.46$ ,  $p = .17$ . The intercepts of the functions were  $.44^\circ$  ( $SD = 7.85$ ) for VM and  $-.55^\circ$  ( $SD = 5.18$ ) for SL; the difference was not significant.

#### Test Phase

*Pointing latency and noise.* Latencies for the two modalities were very similar to those obtained in Experiment 1 (Figure 2). The average latency for VM was significantly shorter than that for SL,  $t(14) = 3.14$ ,  $p < .01$ . Similarly, standard deviations for the two modalities were quite similar to the corresponding values in Experiment 1 (Figure 2). The standard deviation for VM was significantly smaller than that for SL,  $t(14) = -3.03$ ,  $p < .01$ . Table 1 indicates that latency and SD were correlated, as they were in Experiment 1.

*Accuracy.* Figure 5 shows participants' average pointing responses as a function of the correct allocentric direction. Figure 6 presents participants' average reported allocentric distance as a function of correct allocentric distance. Distance standard deviation was significantly lower for VM (.77,  $SD = 0.47$ ) than SL (1.12,  $SD = .65$ ),  $t(14) = -2.31$ ,  $p < .05$ .

*Correlations between modalities.* Despite the difference in means, latency was highly correlated between the two modalities across the target pairs (Table 2). Similarly, despite mean differences, the two modalities correlated significantly in terms of signed error and standard deviation in pointing. These patterns suggest that the introduction of sequential encoding for VM, like that of SL, increased common sources of systematic variation and noise across stimuli, relative to Experiment 1.

### Discussion

Experiment 2 compared reports of allocentric relations while controlling for the sequencing of target presentation in the two modalities. In both the visual memory and spatial language con-

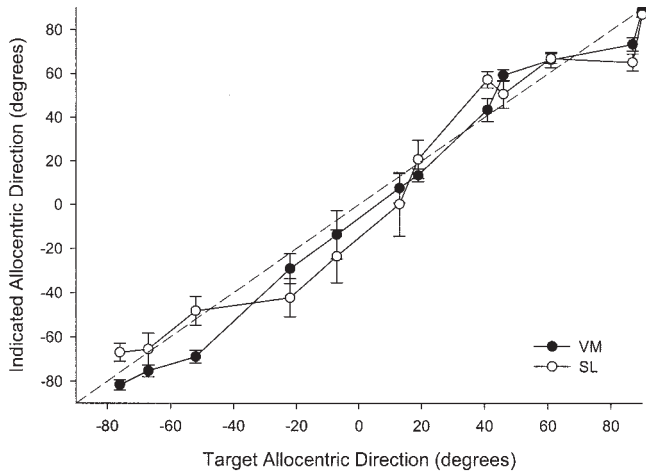


Figure 5. Reported allocentric direction response as a function of physical allocentric directions in Experiment 2. VM = visual memory; SL = spatial language.

ditions, participants learned the locations of targets that were presented to them one after another. Despite this modification, the critical results from Experiment 2 replicated those of Experiment 1. Participants in the VM condition were faster at pointing, and there were corresponding advantages in pointing standard deviation and absolute error. This suggests that the simultaneous presentation of visual targets in Experiment 1 was not the cause of the performance advantage documented with the visual conditions. What then is the cause of this difference between the two modalities?

A hypothesis is that participants in our spatial language condition did not construct a spatial image but instead maintained a verbally based representation of the layout information (e.g., lexical or propositional). Perhaps, then, participants simply remembered the four statements describing the egocentric locations of our targets and formed a spatial image only when allocentric relations were probed. This tendency may have been promoted by the random order in which the targets were presented (see De Vega et al., 2001).

Previous studies (e.g., Klatzky et al., 2003) included spatial language conditions that were identical to the ones run in the present study. However, because in the Klatzky et al. (2003) study participants knew that they had to locate the targets from new standpoints (when they had to walk to them via indirect paths), they could have deemed the strategy of relying on a verbally mediated representation virtually useless. The fact that participants in that study walked immediately and accurately to the targets from the new standpoints suggests that they had formed a spatial image that they continuously updated as they moved. In contrast, the learning phase of Experiments 1 and 2 required only that participants locate individual targets from the standpoint where they had learned them. If indeed our participants did not construct a spatial image from language, perhaps modifying the task to require updating the egocentric information provided at learning could elicit the encoding of an initial egocentric spatial image from which allocentric relations could be computed. To examine this, we inserted an updating task between the learning and the test

phases. That is, after participants had learned the locations of targets egocentrically, they were asked to perform a backward translation and then locate the targets again. To perform the updating task, participants had to convert the statements into spatial images prior to moving. After moving, they performed the allocentric judgments just as in the previous experiments. Note that the backward translation changes egocentric relations but not the allocentric ones. However, the goal here was to induce the formation of the egocentric image. Indeed, in postexperimental interviews, all participants of Experiment 3 reported forming a spatial image of the layouts they had learned.

### Experiment 3

In Experiment 3, after having learned the locations of targets, participants were guided to a novel standpoint that was 90 cm to the back of the original standpoint and were asked to locate the targets from there. Backward movement was chosen to avoid having objects end up being behind the participant (as would happen with forward movement) and ensures that all egocentric directions would change as a result of the movement (targets directly on the left–right would have stayed in the same direction if sideways movement was used instead). Because participants were told before the experiment about this additional task, we expected that they would have chosen to form a spatial image and update it during the backward movement. Previous studies have shown that updating a spatial image during the course of physical movements is performed effortlessly and accurately (e.g., Farrell & Thomson, 1999; Loomis, Klatzky, Golledge, & Philbeck, 1999), particularly with translational body movements. If the reason spatial language was not at par with visual memory in the previous experiment was the absence of a spatial image, then no performance difference between the two modalities was to be expected in this experiment.

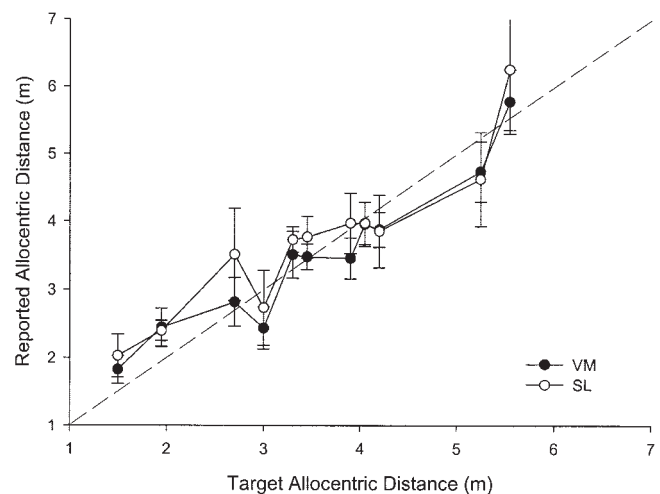


Figure 6. Reported allocentric distance as a function of physical allocentric distance in Experiment 2. VM = visual memory; SL = spatial language.



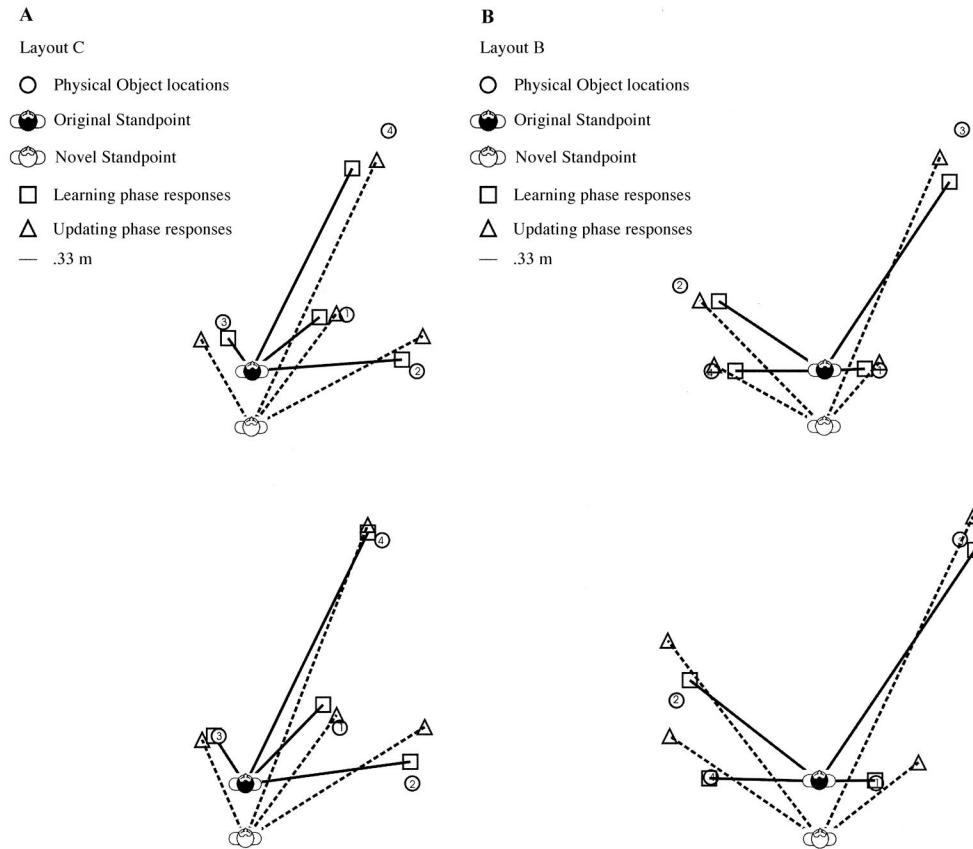


Figure 7. Egocentric responses from the last learning trial and the updating trial in Experiment 3. The top panels in (A) Layout C and (B) Layout B show visual memory; lower panels show spatial language.

### Method

#### Participants

Sixteen (6 male, 10 female) students of introductory psychology classes at the University of California, Santa Barbara, participated in the experiment in exchange for course credit. One additional participant was not accustomed to using feet as units of distance measurement, and another was extremely slow at allocentric pointing (mean RT was longer than 3 standard deviations from the group mean). Both participants were replaced.

#### Design and Procedure

Experiment 3 was identical to Experiment 2 with one exception. After participants had reached the learning criterion they were guided without vision to a new standpoint that was located 0.90 m backward from the original standpoint. From this new standpoint they first performed one egocentric trial with the four objects and then performed the allocentric judgments as in the previous experiments.

### Results<sup>4</sup>

#### Learning Phase

In contrast to the previous experiments, in Experiment 3 participants needed the same number of trials to reach the learning criterion with SL (3.18 trials;  $SD = .54$ ) and VM (3.25 trials;  $SD$

$= .58$ ),  $t(15) = -.29$ ,  $p = .77$ . All trials that failed to reach the learning criterion (three in VM and four in SL) involved categorical errors. Although absolute error for pointing responses was somewhat greater for SL ( $8.04^\circ$ ;  $SD = 4.54$ ) than VM ( $6.36^\circ$ ;  $SD = 2.96$ ), this difference was not statistically reliable,  $t(15) = -1.61$ ,  $p = .13$ .

As in the previous experiments participants' egocentric pointing responses were used to compute psychophysical functions to relate responses to correct values. The slope for VM was .993 ( $SD = .01$ ) and for SL was .988 ( $SD = .02$ ). These values were statistically equivalent,  $t(15) = 1.15$ ,  $p = .27$ . The intercepts of the functions were  $.06^\circ$  ( $SD = 5.46$ ) for VM and  $-.15^\circ$  ( $SD = 6.60$ ) for SL and did not differ significantly,  $t(15) = .09$ ,  $p = .93$ .

#### Updating Phase

Responses collected from the novel standpoint demonstrated the expected pattern; that is, all participants modified their responses from the learning phase to account for the backward translation (Figure 7A, B). Direction responses were moved toward the mid-

<sup>4</sup> Given that in Experiment 3 we expected equivalence, we reported statistical tests for all results, significant or not. For Experiment 3, we conducted prospective power analyses for all comparisons that were significant in Experiment 2. All power estimates had a value of 1.

line, and distances indicated were greater than those in the last trial of the learning phase. This pattern was observed with both the VM and the SL conditions; for direction responses, the average shifts toward the midline were  $17.75^\circ$  ( $SD = 9.94$ ) in the vision condition and  $16.42^\circ$  ( $SD = 7.32$ ) in the spatial language condition,  $t(15) = -.36, p = .73$ . For distances, the estimates given were on average greater than those in the training phase by 0.85 m in the vision condition and 0.98 m in the spatial language condition,  $t(15) = -1.15, p = .27$ .

Furthermore, absolute pointing error, calculated as the absolute difference between the pointing response after updating and the correct updated response, was equal for the two modalities,  $t(15) = .53, p = .60$ ;  $9.86^\circ$  ( $SD = 4.85$ ) and  $8.94^\circ$  ( $SD = 4.34$ ), respectively, for VM and SL. However, with SL, people tended to overestimate the distance of the targets from the new standpoint; signed error for distance estimates was 0.44 m ( $SD = 0.56$ ) for SL and  $-0.16$  m ( $SD = 0.35$ ) for VM;  $t(15) = -5, p < .001$ .

Psychophysical functions that related pointing responses to correct performance were conducted for the updating phase as well. As in the learning phase, the slopes for the two functions were statistically equal,  $t(15) = .16, p = .87$ . For VM the slope was .985 ( $SD = 0.01$ ) and for SL was .984 ( $SD = 0.02$ ). The intercepts of the two functions were  $.05^\circ$  ( $SD = 9.10$ ) for VM and  $-.36^\circ$  ( $SD = 7.43$ ) for SL, which did not differ significantly. Finally, psychophysical functions relating distance estimates to correct updated distances were also conducted. As in the direction functions, the slopes and intercepts of the VM and SL distance functions were equal. For VM the slope was .958 ( $SD = 0.06$ ) and for SL it was .974 ( $SD = 0.05$ ),  $t(15) = -.71, p = .50$ . The intercepts were 0.34 m ( $SD = 0.45$ ) for VM and 0.56 m ( $SD = 0.78$ ) for SL,  $t(15) = -1.02, p = .33$ .

Overall, the results from the updating phase of the experiment indicated that participants were able to update successfully the egocentric locations of targets with both visual memory and spatial language.

### Test Phase<sup>5</sup>

**Pointing latency and noise.** In contrast to both Experiments 1 and 2, latencies for VM and SL were statistically equivalent,  $t(15) = -.24, p = .82$  (Figure 2). Notably, this equivalence came about because of a decrease in the SL latency (5.72 s) from the much higher values, 7.55 and 7.69 s, obtained in Experiments 1 and 2, respectively. In addition, in contrast to the preceding experiments, standard deviation was no greater with VM than with SL,  $t(15) = .08, p = .94$ . Table 1 indicates that pointing latency and standard deviation were correlated, as in the two previous experiments.

**Accuracy.** Figure 8 shows participants' average signed error in pointing responses. A striking similarity across modalities can be seen. Figure 9 presents signed error in participants' reported allocentric distance. Again, the patterns across modalities are strikingly similar. Furthermore, in terms of the distance standard deviation, the difference between VM (.85,  $SD = 0.36$ ) and SL (.91,  $SD = 0.40$ ) was not significant,  $t(15) = -.47, p = .65$ .

**Correlations between modalities.** As shown in Table 2, latency was significantly correlated between VM and SL, as were the pointing signed error and pointing standard deviation. Furthermore, VM and SL correlated significantly in terms of distance

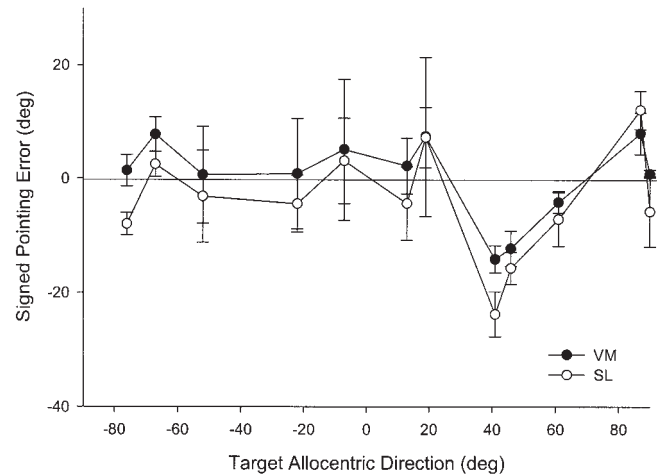


Figure 8. Signed pointing error as a function of physical allocentric direction in Experiment 3. deg = degrees; VM = visual memory; SL = spatial language.

signed error, and there was a similar trend in distance standard deviation ( $r = .42, p < .20$ ). It should be noted that the correlations in signed error were very high despite the fact that the paired values came from different participants.

**Scaling VM and SL responses.** Finally, we constructed for one layout (Layout C) two-dimensional scaling solutions for the data in Experiment 3, separately for each modality and for each response type (angles and distances, as suggested by Waller & Haun, 2003). This was done by assigning Cartesian coordinates to the target locations that minimized the total absolute differences between the scaled value and the response values (angle or distance). The location of 1 point was fixed as an origin. The four solutions were placed in a common coordinate system and aligned with respect to the bearing from Point 1 to Point 4, which was set to the objective bearing. The solutions were also scaled; thus, the average distance equaled the average objective distance. (Alternatively, one could scale and rotate each solution relative to the average responses, but this would make little difference because the average errors were small: The distance error was  $-0.10$  m for SL and  $-0.05$  m for VM, and the angle error was  $-4.0^\circ$  for SL and  $0.5^\circ$  for VM.) These scaling solutions are shown in Figure 10 and indicate an obvious correspondence across modalities and response measures.

### Discussion

Experiment 3 required that participants update the egocentric relations that were provided during the learning phase. Although this updating requirement does not affect allocentric relations, the results from this experiment differed from those of the first two experiments. In contrast to both Experiments 1 and 2, participants in Experiment 3 were equally fast and variable with spatial lan-

<sup>5</sup> One VM and two SL observations with pointing responses deviating more than  $60^\circ$  from the correct direction were considered outliers, and both angles and distances were discarded from all analyses. The same criterion was used in both Experiments 1 and 2, but no outliers were found in those experiments.

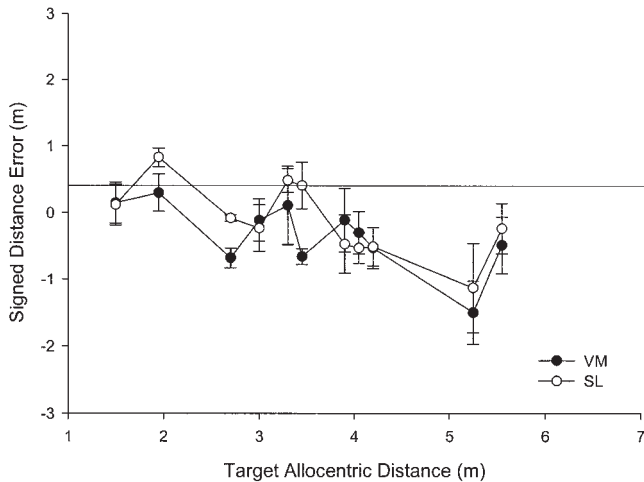


Figure 9. Signed distance error as a function of physical allocentric distance in Experiment 3. VM = visual memory; SL = spatial language.

guage and vision. Moreover, the latencies, errors, and noise measures tended to be highly correlated between the two modalities. Finally, the scaling solutions that were done separately for each modality and response type clearly agreed and were close to the locations of the targets.

We believe that the updating requirement encouraged participants to construct an egocentric spatial image, which was functionally equivalent across the two modalities, from which allocentric relations were computed. The results of Experiment 3 support both the image-updating model of Loomis et al. (2002) and Bryant's (1997) spatial representation system.

### General Discussion

The present experiments investigated whether spatial representations derived from vision and spatial language are equivalent in terms of enabling the report of allocentric spatial relations. The experiments began with a three-phase model of the task, which assumes vision and language differ in the processes that encode spatial images, but once the images have been formed, computing and reporting allocentric relations involves common processes. The first two experiments we conducted produced results at odds with the functional equivalence hypothesis. In both experiments, performance on allocentric spatial relations was faster and less variable following visual encoding than with spatial language, even when the responses in the visual condition were made from memory. However, we believe that the disadvantage with spatial language was due to the use of a different strategy in the spatial language condition. Participants in Experiments 1 and 2 had no reason to convert the verbal descriptions that provided the egocentric relations into a spatial image until the time of the allocentric test. We believe that participants who were given linguistic descriptions used a verbally based representation during the initial learning phase, deferring the formation of a spatial image until the allocentric relations were tested. This deferred processing caused them to require longer latencies when reporting allocentric directions and produced higher noise.

An interesting question that arose is whether the formation of the spatial image in the SL condition of Experiments 1 and 2

occurred in a piecemeal fashion (i.e., participants formed a spatial image containing the locations of only the current target pair) or in an all-at-once fashion during the first trial of the test phase. Additional statistical analyses showed that, in all three experiments, the changes in the patterns of results for the various dependent measures were negligible if the first trial of the test phase was excluded. This finding is consistent with a piecemeal account of image formation.

A related question is whether allocentric relations, in the vision conditions, were computed from an egocentric spatial image at the time of test or whether an allocentric spatial image was constructed during learning. A recent theory of spatial memory by McNamara and colleagues (e.g., McNamara, 2003; Mou & McNamara, 2002; Shelton & McNamara, 2001) proposes the latter possibility. According to this theory, interobject relations are represented in a spatial image organized in terms of an intrinsic reference frame chosen on the basis of various cues (e.g., egocentric viewpoint, instructions, characteristics of the layout). Our data do not speak to the question of whether allocentric relations were encoded during learning or at test, except for the conditions in which formation of a spatial image was deferred (SL in Experiments 1 and 2), which would correspondingly defer extracting allocentric relations.

Another important finding coming from additional analyses we performed is that participants who had SL as the second or a third condition in the testing order showed the same pattern of results as the ones who had it first. This suggests that prior experience with the task did not probe them to spontaneously form a spatial image in the SL condition in Experiments 1 and 2. We believe that this was the case because participants in the first two experiments knew that they could rely on their memories for the egocentric relations and could defer image formation and allocentric computation.

Results were different in Experiment 3, in which an updating requirement was added. Performance in the spatial language condition, as measured in terms of standard deviation and latency, was at par with that in the visual-memory condition. In addition, the systematic errors (Figures 8 and 9) were remarkably similar for the two modalities. We believe that the updating requirement in Experiment 3 encouraged participants to convert the verbal descriptions to a spatial image. Results from the learning phase of Experiment 3 suggest that the crucial factor might not have been the

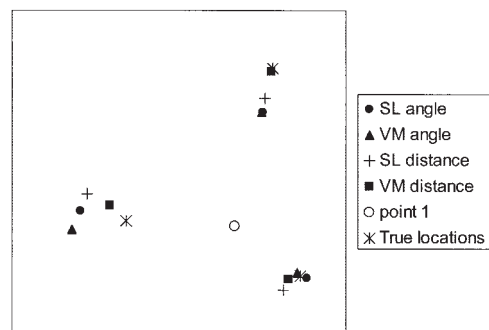


Figure 10. Scaling of Layout 1 in Experiment 3 by four sets of data: spatial language (SL) angle responses, visual memory (VM) angle responses, SL distance responses, and VM distance responses. Objective target locations are also shown.

updating act itself but the expectation of it. In contrast to the first two experiments, VM and SL did not differ in terms of any of the measures of the learning phase. This suggests that functional equivalence was present early in the experiment.

In keeping with Loomis et al. (2002) and Bryant (1997), we suggest that the spatial images constructed from spatial language and vision were functionally equivalent. The results from Experiment 3 therefore extend previous findings by showing that spatial language can produce spatial images from which participants can easily extract information that was not explicitly learned. However, note that the present study and the studies by Loomis et al. (2002) and Klatzky et al. (2003) used rather simple spatial layouts. Therefore, it is still possible that functional equivalence will not be evidenced with tasks involving more complex layouts or tasks with different requirements (e.g., tasks that depend on the visual richness of the scene).

Specific predictions in regard to functional equivalence of VM and SL were upheld in Experiment 3. The allocentric report latencies were equal, and the two modalities were then also equal with respect to noise. When computed across target pairs, there were strong correlations between the modalities with respect to signed error, noise, and latency. As a final point, the scaling solution for each modality and response type (i.e., direction or distance) were matched closely to one another and to the locations of the targets, supporting the idea of a common representation underlying performance.

It is directly predicted by the model in Figure 1 that if participants delay in encoding language into a spatial image, their latency to make allocentric reports will increase. The question arises as to why the delay introduces additional noise (increase in standard deviation) as well as reducing the correlations between the systematic errors for SL and VM, correlations that were quite high in Experiment 3. One possibility is that in delaying the transition to the allocentric task, SL participants forgot some of the verbal information they had just learned. Such a memory-based effect is consistent with the finding of Spencer and Hund (2002), that both systematic errors and variability in pointing to previously seen targets on a table top increased with delays of up to 20 s between presentation and test. Another possibility is that participants remembered the information but terminated the image-encoding process earlier after deferring it to the test phase than if they had performed it in the learning phase. A third possibility is that subjects created a separate spatial image of each object in the learning phase and then tried to integrate the images at the time of test. In any case, the data indicate that the deferred encoding for spatial language in the first two experiments introduced not only a longer latency to report the allocentric relations, but also introduced noise and diminished the correlated pattern of systematic error.

An interesting and unexpected result to come of this work is the high correlation between noise and latency (Prediction 5). The correlation between the seven pairs of values (latency and standard deviation) in Figure 2 is .91. Even within each modality in the three experiments (Table 1), the correlations were all positive, averaging .59 for the seven values. A possible way of understanding why latency and noise might be correlated is that observed noise is correlated with the participants' uncertainty (judgmental noise) and that this uncertainty is highly correlated with the re-

sponse latency, which is often used as a measure of memory strength. Because the observed noise includes variability associated with making the pointing response, this line of thinking suggests that the correlation between latency and noise would be even higher if it were possible to measure judgmental noise directly.

A final intriguing issue is what brain areas mediate the functional equivalence across modalities that we have found for egocentric spatial updating and allocentric judgments. Anatomical and information-processing characteristics point to the posterior parietal cortex (PPC) as playing a principal role in these tasks. The PPC receives inputs from visual, auditory, and somatosensory sources. It maps spatial representations between different reference frames, for example, as defined by the eye, head, or body (Cohen & Andersen, 2002; Snyder, Grieve, Brotchle, & Andersen, 1998). PPC has been associated with visual memory (Sereno, Pitzales, & Martinez, 2001) and imagery requiring spatial manipulation (Farah, Hammond, Levine, & Calvanio, 1988; Kosslyn & Thompson, 2003). Hippocampal areas may also be involved in the allocentric judgments tested in the present studies. Burgess and his colleagues (Burgess et al. 1999; Burgess, 2002) suggested that PPC may function particularly for forming egocentric representations, providing inputs to the hippocampus for use in deriving allocentric maps.

Spatial imagery in PPC is to be distinguished from specifically visual imagery, which Kosslyn and colleagues called depictive (Kosslyn, 1994; Kosslyn, Ganis, & Thompson, 2003). By depictiveness it is meant that there is a direct correspondence between regions of the represented object and those of the image, such that interpoint distance is preserved. Kosslyn and colleagues (Kosslyn & Thompson, 2003; Kosslyn et al., 2003) proposed that depictive imagery corresponds to activation in specifically visual areas of the brain, including V1, that provide fine spatial resolution.

The results from the present experiments have importance for the design of auditory displays of spatial information. For example, there has been great interest in developing GPS-based navigation systems for blind people, and the question is how best to display information to blind travelers to guide them along routes and to inform them of points of interest in the environment. The more typical approach is to use synthesized speech to convey information about routes and points of interest (e.g., Helal, Moore, & Ramachandran, 2001; LaPierre, 1998; Makino, Ishii, & Nakashizuka, 1996; Petrie, Johnson, Strothotte, Raab, Fritz, et al., 1996).

An alternative, however, is to use some form of spatialized display that conveys spatial information more directly through hearing or touch (e.g., Golledge, Klatzky, Loomis, Speigle, & Tietz, 1998; Loomis, Golledge, Klatzky, Speigle, & Tietz, 1994). Spatialized sound, for example, is somewhat better than spatial language in guiding people over a route (Loomis, Golledge, & Klatzky, 2001) and in the rapidity with which people can learn a layout of landmarks, as shown by results of Klatzky et al. (2002, 2003). However, the effectiveness of spatial language in building up spatial images suitable for spatial updating (Loomis et al., 2002; Klatzky et al., 2003) and for allocentric judgments, as shown by Experiment 3, indicates that spatial language is an effective alternative as a way of displaying spatial information.

## References

- Avraamides, M. N. (2003). Spatial updating of environments described in texts. *Cognitive Psychology*, *47*, 402–431.
- Bryant, D. J. (1997). Representing space in language and perception. *Mind & Language*, *12*, 239–264.
- Bryant, D. J., Lanca, M., & Tversky, B. (1995). Spatial concepts and perception of physical and diagrammed scenes. *Perceptual and Motor Skills*, *81*, 531–546.
- Bryant, D. J., & Tversky, B. (1999). Mental representations of perspective and spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 137–156.
- Burgess, N. (2002). The hippocampus, space, and viewpoints in episodic memory. *The Quarterly Journal of Experimental Psychology*, *55A*, 1057–1080.
- Burgess, N., Jeffery, K., & O'Keefe, N. (1999). Integrating hippocampal and parietal functions: A spatial point of view. In N. Burgess, K. J. Jeffery, & J. O'Keefe (Eds.), *The hippocampal and parietal foundations of spatial cognition* (pp. 3–29). London: Oxford University Press.
- Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews: Neuroscience*, *3*, 553–562.
- De Vega, M., Cocude, M., Denis, M., Rodrigo, M. J., & Zimmer, H. D. (2001). The interface between language and visuo-spatial representations. In M. Denis, R. H. Logie, C. Cornoldi, M. De Vega, & J. Engelkamp (Eds.), *Imagery, language, and visuo-spatial thinking* (pp. 109–136). Hove, England: Psychology Press.
- De Vega, M., & Rodrigo, M. J. (2001). Updating spatial layouts mediated by pointing and labelling under physical and imaginary rotation. *European Journal of Cognitive Psychology*, *13*, 369–393.
- Denis, M., & Cocude, M. (1989). Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology*, *1*, 293–307.
- Denis, M., & Cocude, M. (1997). On the metric properties of visual images generated from verbal descriptions: Evidence for the robustness of the mental scanning effect. *European Journal of Cognitive Psychology*, *9*, 353–379.
- Denis, M., & Zimmer, H. D. (1992). Analog properties of cognitive maps constructed from verbal descriptions. *Psychological Research*, *54*, 286–298.
- Elliott, D., & Madalena, J. (1987). The influence of premovement visual information on manual aiming. *Quarterly Journal of Experimental Psychology*, *A39*, 541–559.
- Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, *20*, 439–462.
- Farrell, J. J., & Thomson, J. A. (1999). On-line updating of spatial information during locomotion without vision. *Journal of Motor Behavior*, *31*, 37–53.
- Federico, T., & Franklin, N. (1997). Long-term spatial representations from pictorial and textual input. In S. C. Hirtle & A. U. Frank (Eds.), *Spatial information theory: A theoretical basis for GIS, Lecture notes in computer science No. 1329*. Heidelberg, Germany: Springer-Verlag.
- Ferguson, E. L., & Hegarty, M. (1994). Properties of cognitive maps constructed from text. *Memory & Cognition*, *22*, 455–473.
- Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, *119*, 63–76.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: Bradford Books/MIT Press.
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*, 69–83.
- Golledge, R. G., Klatzky, R. L., Loomis, J. L., Speigle, J., & Tietz, J. (1998). A geographic information system for a GPS based personal guidance system. *International Journal of Geographical Information Science*, *12*, 727–749.
- Helal, A., Moore, S., & Ramachandran, B. (2001, October). *Drishii: An integrated navigation system for visually impaired and disabled*. Proceedings of the 5th International Symposium on Wearable Computers, Zurich, Switzerland.
- Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology*, *13*, 149–206.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, England: Cambridge University Press.
- Johnson-Laird, P. N. (1996). Space to think. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 437–462). Cambridge, MA: MIT Press.
- Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freska, C. Habel, & K. F. Wender (Eds.), *Spatial cognition—An interdisciplinary approach to representation and processing of spatial knowledge (Lecture notes in artificial intelligence 1404)* (pp. 1–17). Berlin, Germany: Springer-Verlag.
- Klatzky, R. L., Lippa, Y., Loomis, J. M., & Golledge, R. G. (2002). Learning directions of objects specified by vision, spatial audition, or auditory spatial language. *Learning and Memory*, *9*, 364–367.
- Klatzky, R. L., Lippa, Y., Loomis, J. M., & Golledge, R. G. (2003). Encoding, learning and spatial updating of multiple object locations specified by 3-D sound, spatial language, and vision. *Experimental Brain Research*, *149*, 48–61.
- Kosslyn, S. M. (1994). *Image and brain*. Boston: MIT Press.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2003). Mental imagery: Against the nihilistic hypothesis. *Trends in Cognitive Sciences*, *7*, 109–110.
- Kosslyn, S. M., Reiser, B. J., Farah, M. J., & Fliegel, S. L. (1983). Generating visual images: Units and relations. *Journal of Experimental Psychology: General*, *112*, 278–303.
- Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, *129*, 723–746.
- LaPierre, C. (1998). *Personal navigation system for the visually impaired*. Unpublished master's thesis, Department of Electronics, Carleton University, Ottawa, Ontario, Canada.
- Loomis, J. M., Golledge, R. G., & Klatzky, R. L. (2001). GPS-based navigation systems for the visually impaired. In W. Barfield & T. Caudell (Eds.), *Fundamentals of wearable computers and augmented reality* (pp. 429–446). Mahwah, NJ: Erlbaum.
- Loomis, J. M., Golledge, R. G., Klatzky, R. L., Speigle, J. M., & Tietz, J. (1994). Personal guidance system for the visually impaired. *Proceedings of the First Annual ACM/SIGGRAPH Conference on Assistive Technologies* (pp. 85–91). New York: Association for Computing Machinery.
- Loomis, J. M., Klatzky, R. L., Golledge, R. G., & Philbeck, J. W. (1999). Human navigation by path integration. In R. G. Golledge (Ed.), *Way-finding: Cognitive mapping and other spatial processes* (pp. 125–151). Baltimore: Johns Hopkins.
- Loomis, J. M., Lippa, Y., Klatzky, R. L., & Golledge, R. G. (2002). Spatial updating of locations specified by 3-D sound and spatial language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 335–345.
- Makino, H., Ishii, I., & Nakashizuka, M. (1996, October–November). *Development of navigation system for the blind using GPS and mobile phone connection*. Proceedings of the 18th Annual Meeting of the IEEE EMBS, Amsterdam, the Netherlands.
- McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In C. Freksa, W. Brauer, C. Habel, & K. Wender (Eds.), *Spatial cognition III: Routes and navigation, human memory and learning, spatial representation and spatial reasoning* (pp. 174–191). Berlin, Germany: Springer-Verlag.

- McNamara, T. P., & Diwadkar, V. A. (1997). Symmetry and asymmetry of human spatial memory. *Cognitive Psychology, 18*, 87–121.
- Mellet, E., Bricogne, S., Crivello, F., Mazoyer, B., Denis, M., & Tzourio-Mazoyer, N. (2002). Neural basis of mental scanning of a topographic representation built from a text. *Cerebral Cortex, 12*, 1322–1330.
- Mellet, E., Bricogne, 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.
- Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26*, 165–187.
- Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 162–170.
- Newcombe, N., Huttenlocher, J., Sandberg, E., Lie, E., & Johnson, S. (1999). What do misestimations and asymmetries in spatial judgment indicate about spatial representation? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 4*, 986–996.
- Petrie, H., Johnson, V., Strothotte, T., Raab, A., Fritz, S., & Michel, R. (1996). MoBIC: Designing a travel aid for blind and elderly people. *Journal of Navigation, 49*, 45–52.
- Richardson, A. E., Montello, D., & Hegarty, M. (1999). Spatial knowledge acquisition from maps, and from navigation in real and virtual environments. *Memory & Cognition, 27*, 741–750.
- Rieser, J. J. (1989). Access to knowledge of spatial structure from novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1157–1165.
- 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.
- Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43*, 274–310.
- Snyder, L. H., Grieve, K. L., Brotchle, P., & Andersen, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature, 394*, 887–890.
- Spencer, J. P., & Hund, A. M. (2002). Prototypes and particulars: Spatial categories are formed using geometric and experience-dependent information. *Journal of Experimental Psychology: General, 131*, 16–37.
- Taylor, H., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language, 31*, 261–282.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Waller, D., & Haun, D. B. (2003). Scaling techniques for modeling directional knowledge. *Behavior Research Methods, Instruments, & Computers, 35*, 285–293.
- Westwood, D. A., Heath, M., & Roy, E. A. (2003). No evidence for accurate visuomotor memory: Systematic and variable error in memory-guided reaching. *Journal of Motor Behavior, 25*, 127–134.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language and memory. *Psychological Bulletin, 123*, 162–185.

Received October 13, 2003

Revision received December 18, 2003

Accepted December 19, 2003 ■

## Low Publication Prices for APA Members and Affiliates

**Keeping you up-to-date.** All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential resources.** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

**Other benefits of membership.** Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

**More information.** Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.