

Object location memory: The interplay of multiple representations

HONGBIN WANG, TODD R. JOHNSON, YANLONG SUN, and JIAJIE ZHANG
University of Texas Health Science Center, Houston, Texas

This article reports three experiments in which the representational nature of human object location memory was studied. The results show that multiple frames of reference can be used to encode the spatial relationships among objects. Depending on their dominancy, availability, and validity, these multiple representations interact to determine memory performance. Specifically, representations that are automatically encoded and extensively practiced are more dominant, and their availability improves performance when they are valid. On the other hand, when the dominant representations are available but invalid, people may have to resort to the less dominant representations. As a result, the availability of these dominant but invalid representations can actually hurt performance, due to interference. If these interfering representations are eliminated, the performance is again improved. The implications of these findings for general human spatial cognition are discussed.

How object locations are represented in memory is an essential issue in the study of human spatial cognition. Although it has been generally agreed that different frames of reference can be used to represent object locations, one central debate focuses on which frame of reference people actually use in various spatial tasks (for some recent reviews, see Burgess, Jeffery, & O'Keefe, 1999a; Lansdale, 1998; Newcombe, 2002; Newcombe & Huttenlocher, 2000; Tversky, 2000; R. F. Wang & Spelke, 2002).

Although different taxonomies exist (e.g., Brewer & Pears, 1993; Garnham, 1989; Klatzky, 1998; Levinson, 1996; Logan & Sadler, 1996; Palmer, 2003; Talmy, 1983), one common approach distinguishes between egocentric and allocentric spatial representations. Following the seminal work of Tolman (1948) and O'Keefe and Nadel (1978), some researchers believe that people represent spatial locations in terms of allocentric-based cognitive maps (see, e.g., Gallistel, 1990; King, Burgess, Hartley, Vargha-Khadem, & O'Keefe, 2002; O'Keefe, 1993, 1996). However, some other researchers have argued that people primarily maintain egocentric-based representations. For example, Diwadkar and McNamara (1997) found that when participants studied a scene from one viewpoint and did a recognition test from a second viewpoint, the reaction time (RT) varied linearly with the angular difference between the views. Similar results along this line have been

reported (e.g., Easton & Sholl, 1995; Franklin, Tversky, & Coon, 1992; Mou & McNamara, 2002; Shelton & McNamara, 1997; Sholl & Bartels, 2002; Sholl & Nolin, 1997; Tversky, 1996; Valiquette, McNamara, & Smith, 2003; Waller, Montello, Richardson, & Hegarty, 2002; R. F. Wang & Simons, 1999; R. F. Wang & Spelke, 2000).

Although the debate continues (for recent reviews, see Driver & Pouget, 2000; McNamara, 2003; R. F. Wang & Spelke, 2002), an integrated view is to assume that people simultaneously maintain multiple different representations in spatial memory. One can, for example, encode the location of a specific coffee cup in a number of different ways. Egocentrically, one may say that "it is in front of me." Allocentrically, one can say that "it is in my office," "it is on the desk in my office," or "it is next to my computer on the desk in my office." Depending on the context and various constraints, people may choose to use different representations for different tasks.

Although the claim that people may simultaneously maintain multiple representations has received an increasing amount of support (e.g., Burgess et al., 1999a; Colby & Goldberg, 1999; Halligan, Fink, Marshall, & Vallar, 2003; McNamara, 2003; Sun, Wang, & Johnson, 2004; H. Wang, Johnson, & Zhang, 2001), how different representations work together to give rise to spatial performance remains unclear. This issue is important in the sense that although mathematically equivalent, different representations with different frames of reference possess different properties and readily support different actions (Klatzky, 1998). When multiple types of representations are simultaneously available, it is likely that it is their interaction, but not any single type, that determines performance.

In this article, we report three experiments that were designed to explicitly address this issue. Specifically, in our experiments, a task was used in which the locations of objects on a computer screen could be encoded in multiple ways. Since different types of representations might be

This research was supported by Grant N00014-01-1-0074 from the Office of Naval Research Cognitive Science Program. We thank Yue Wang and Ruijun Bao for their assistance in data collection and Jin Fan for providing the stimulus database. We also thank the action editor and three anonymous reviewers for their critical comments and suggestions that helped to improve the quality of this article. Correspondence concerning this article should be sent to H. Wang, School of Health Information Sciences, University of Texas Health Science Center at Houston, 7000 Fannin, Suite 600, Houston, TX 77030 (e-mail: hongbin.wang@uth.tmc.edu).

simultaneously available and applicable for later problem solving, this task allowed us to explicitly examine the possible interplay of different representations. Although our experiments focus only on various allocentric representations, the results support the general claim that multiple frames of reference can be used to encode spatial relations among objects. In addition, we suggest that when multiple representations are available for spatial problem solving, they interact to determine performance.

The Experimental Paradigm

We adopted an experimental paradigm developed by Milner and colleagues in the 1990s, which we will call the *Milner paradigm* (e.g., Johnsrude, Owen, Crane, Milner, & Evans, 1999; Milner, Johnsrude, & Crane, 1997; Owen, Milner, Petrides, & Evans, 1996). Although their focus was on using neuroimaging techniques to probe the

brain foundations of object location memory, the Milner paradigm offers an experimental design that allows a systematic evaluation of multiple schemas for representing spatial relationships.

The paradigm examines human object location memory, using the standard study–test methodology. In the study phase, eight *objects* (black-and-white *drawings*) are presented, one at a time, on a computer screen, with each accompanied by two solid black squares as *landmarks*.¹ Participants are asked to study and remember the locations of these objects, relative to the landmarks. As an example (see Figure 1A), suppose that a *chair* is one of the eight objects that the participants study. Furthermore, assume that the location of the chair relative to the landmarks is encoded as *loc-chair*. In the testing phase, the participants are presented with a pair of identical studied objects (say, two chairs), together with some retrieval

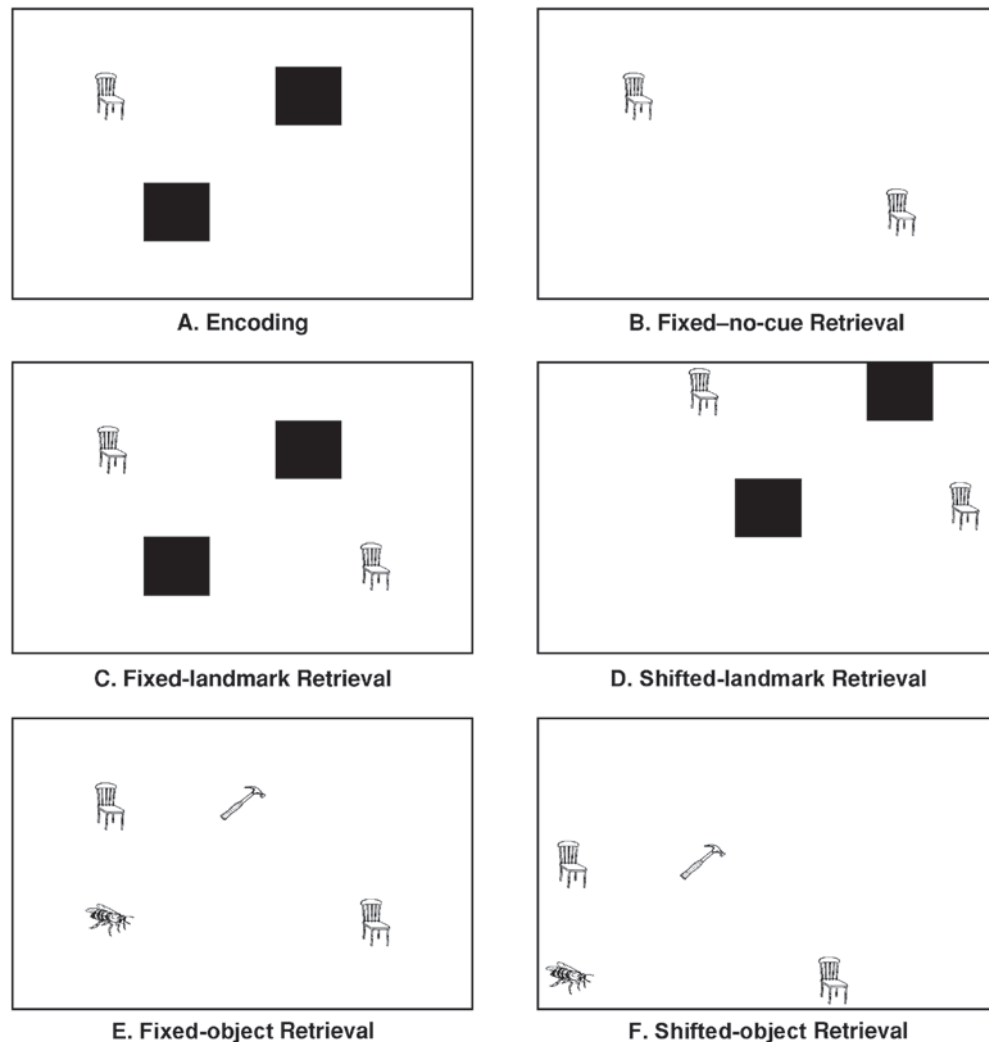


Figure 1. The design of Experiment 1. Panel A represents a study trial (i.e., remember the location of a chair relative to the two landmarks), and panels B–F represent various testing trials (i.e., identify one of the two chairs that is in its original studied location, relative to the screen, landmarks, or other studied objects).

cues. The two identical objects are presented at different locations, with one, which we call the *target*, in its original studied location, relative to the retrieval cues (i.e., location *loc-chair*), and the other one, which we call the *distractor*, in a different location (i.e., not in location *loc-chair*). The participants are required to perform a forced choice task to select the target out of the target–distractor pair.

By manipulating the retrieval cues and whether the spatial array on the screen is fixed or shifted, Milner and colleagues (e.g., Johnsrude et al., 1999; Milner et al., 1997) generated and studied the following four testing conditions:

1. In the *fixed-landmark* condition (Figure 1C), the two landmarks are presented as the retrieval cues, along with the target–distractor pair (e.g., two chairs). The absolute locations of landmarks and objects on the screen are unchanged from their original studied positions.

2. In the *shifted-landmark* condition (Figure 1D), the two landmarks are presented as the retrieval cues, along with the target–distractor pair (e.g., two chairs). Although the relative spatial relationships among the landmarks and objects remain unchanged, the whole array of landmarks and objects on the screen is shifted. The shift can occur in eight possible directions, including up–left, left, down, and so forth, as long as the landmarks and objects are all visible and their original spatial relationships remain.

3. In the *fixed-object* condition (Figure 1E), two randomly selected studied objects (e.g., a hammer and a bee), instead of the two landmarks, are presented as the retrieval cues, along with the target–distractor pair (e.g., two chairs). The absolute locations of objects on the screen are unchanged. In order to correctly identify the target in this condition, presumably, relevant object-to-object relationships are necessary. Since objects are never co-presented in the study phase, their spatial relationships may not be readily available and may have to be computed through a chain of inferences (e.g., from the locations of two object cues, to infer the locations of landmarks and then to infer the location of target).

4. In the *shifted-object* condition (Figure 1F), two randomly selected study objects (e.g., a hammer and a bee), instead of the two landmarks, are presented as the retrieval cues, along with the target–distractor pair (e.g., two chairs). This condition is similar to the fixed-object condition, except that the whole array of objects on the screen is now shifted. The shift occurs in a fashion similar to that in the shifted-landmark condition. However, since the objects' locations relative to each other are unchanged, presumably, the participants can still identify the target through a process similar to the one in the fixed-object condition.

One important feature of the Milner paradigm is that it simultaneously involves multiple types of spatial representations for object locations. In this study, we focus only on those allocentric representations, including screen-based (the location of an object relative to the computer screen), landmark-based (the location of an object relative to the landmarks), and object-based (the location of an object relative to other objects) ones. As we will demonstrate in

our experiments, it seems that screen-based representations can be automatically encoded and maintained in the study phase, due to the steady presence of the computer screen (see Naveh-Benjamin, 1987, 1988). In addition, since an object is always presented along with the two landmarks in the study phase and participants are explicitly instructed to encode the locations of objects relative to the landmarks, landmark-based representations can also be constructed in the study phase. However, object-based representations are different. Although object-to-object relationships are kept fixed in the study phase, no two objects are ever presented on the screen at the same time. As a result, these relationships may not be readily encoded in the study phase and have to be computed when needed later in the testing phase. The systematic manipulation of the different testing conditions allowed Milner and colleagues (e.g., Johnsrude et al., 1999; Milner et al., 1997) to use a subtraction method to determine the brain areas that were activated in the different test conditions. Behavioral data were only briefly reported in Johnsrude et al. (1999). It was found that the shifted-object condition was harder (e.g., longer RTs and lower accuracies) than any other conditions, which did not differ from each other.

In sum, the Milner paradigm offers an elegant framework for studying the representational nature of human object location memory, in that multiple forms of spatial representations with different frames of reference may be simultaneously involved. Although interesting results have been obtained, important details, such as what spatial representations underlie each testing condition and how different spatial representations may interact with each other to give rise to performance, remain unclear. In the experiments we report next, we extended the Milner paradigm to make the distinctions among different spatial representations more explicit. In doing so, we expected that the possible interaction among these representations could be examined in more detail and that a better understanding of human object location memory could be obtained.

EXPERIMENT 1

The main purpose of Experiment 1 was to examine the underlying spatial representations and their interaction in each testing condition in the Milner paradigm by manipulating retrieval cues and fixed/shifted array types. To this end, we added one more testing condition to the original Milner paradigm. In this additional condition, called the *fixed-no-cue* condition, no retrieval cues were presented with the target–distractor pair in the testing phase (see Figure 1B). The participants had to make the forced choice solely on the basis of the absolute locations of objects on the screen. This condition was added to explicitly test the effect of screen-based spatial representations in location retrieval.

Method

Participants. Twenty-one college students from the University of Texas Health Science Center at Houston, with normal or corrected-to-normal vision, were paid to participate in the experiment.

Apparatus and Materials. Five sets of stimuli (each consisting of eight black line drawings of common objects) were selected from the database developed by Snodgrass and Vanderwart (1980; see the Appendix). The object drawings, 100×100 pixels in size, were presented against a white background on a 19-in. VGA monitor with a resolution of $1,024 \times 768$. The monitor was in front of the participants, within 2 ft. The participants were asked to respond by clicking with a mouse that was within comfortable reach.

Design and Procedure. The experiment adopted a within-subjects design. Each participant performed in five sessions, with each consisting of one study phase and one testing phase. Whereas the study phase was identical in all five sessions, a different testing condition, as depicted in Figures 1B–1F, was used in the testing phase of each session. The session order for each participant was randomized. A different stimulus set was used in each session.

In each trial of the study phase, the participants were presented with one object (a drawing), together with two landmarks, and were instructed to remember the location of the object relative to the landmarks. The participants controlled the pace of the experiment by using the mouse to click on the object whenever they were ready to move on to the next study trial. There were 32 study trials in each session, with each object presented four times. The presentation order was randomized. During the study phase, the participants did not know which testing condition would follow.

After the study phase, the participants went on to a testing phase. Each testing phase consisted of a block of 32 testing trials of the same testing condition. Each of the eight encoded objects was tested four times. In each testing trial, the participants were presented with a target–distractor pair, together with the corresponding retrieval cues according to the testing condition of the session (see Figures 1B–1F). The participants were instructed to identify the target and respond by clicking it (using the mouse) as quickly and accurately as possible. As soon as the participants clicked, the next testing trial was presented. The RT and accuracy data were recorded.

Results

Accuracy data. The average accuracy for the five testing conditions ranged from 93.8% to 97.8% (see Table 1), and there were no differences among the conditions [$\chi^2(4) = 0.81, p > .93$].

RT data. The RT data from all the participants were pooled together, and those trials with (1) wrong answers or (2) RTs that were three standard deviations above the

corresponding condition means were designated as invalid (about 8.8% of the total trials) and were eliminated from further analyses.

The valid RT data for all five testing conditions are shown in Figure 2 (see Table 1 for numerical values). The general pattern of the data was largely consistent with Johnsrude et al.'s (1999) results, showing that the object-based conditions were harder than the landmark-based conditions and the shifted arrays were harder than the fixed arrays. A repeated measures ANOVA with a 2 (retrieval cue type: landmark vs. object) \times 2 (array type: fixed vs. shifted) design showed a significant interaction between retrieval cue type and array type [$F(1,20) = 24.46, p < .001$]. The main effects of both cue type and array type were significant [$F(1,20) = 43.56, p < .001$, and $F(1,20) = 30.04, p < .001$, respectively]. An additional repeated measures ANOVA on the five testing conditions showed a significant effect of the testing conditions [$F(4,80) = 66.25, p < .001$]. Multiple planned paired *t* tests were then conducted to further test the simple effects among different testing conditions. All the tests were corrected on the basis of Keppel's (1991) modified Bonferroni test, and an adjusted significance level of .02 was adopted.² The results revealed that although there was no significant difference between the fixed–no-cue and the fixed–landmark conditions (RT difference = 86 msec, $p > .10$), both conditions had significantly shorter RTs than did the other three conditions [minimal RT difference = 378 msec; $t(20) = -2.72, p < .014$]. In addition, the shifted–landmark condition had shorter RTs than did the shifted–object condition [RT difference = 1,836 msec; $t(20) = -8.01, p < .001$], and the fixed–object condition had shorter RTs than did the shifted–object condition [RT difference = 1,558 msec; $t(20) = -7.11, p < .001$].

Discussion

Experiment 1 revealed three major findings that have not been reported by Milner and colleagues. First, partici-

Table 1
The Representational Decomposition for the Experiments

| Testing Condition | Experiment | RT \pm SD (ms) | Accuracy (%) | Screen-Based Representations | Landmark-Based Representations | Object-Based Representations |
|-------------------|------------|-------------------|--------------|------------------------------|--------------------------------|------------------------------|
| Fixed, no cue | 1 | 1,592 \pm 270 | 95.4 | + | | |
| | 2 | | | | | |
| | 3 | | | | | |
| Fixed landmark | 1 | 1,506 \pm 253 | 97.8 | + | + | |
| | 2 | 1,258 \pm 311 | 99.5 | | + | |
| | 3 | 1,552 \pm 507 | 95.8 | + | + | |
| Shifted landmark | 1 | 1,970 \pm 585 | 94.3 | – | + | |
| | 2 | 1,511 \pm 310 | 93.2 | | + | |
| | 3 | 1,827 \pm 557 | 93.3 | – | + | |
| Fixed object | 1 | 2,247 \pm 325 | 96.6 | + | | ⊕ |
| | 2 | 2,135 \pm 322 | 92.7 | | | ⊕ |
| | 3 | 2,117 \pm 601 | 92.4 | + | | ⊕ |
| Shifted object | 1 | 3,806 \pm 1,151 | 93.8 | – | | ⊕ |
| | 2 | 3,619 \pm 966 | 88.8 | | | ⊕ |
| | 3 | 3,055 \pm 1,009 | 88.4 | – | | ⊕ |

Note—Different symbols are used to indicate the availability and the applicability of different representations (columns) in different testing conditions (rows). In particular, + designates available and applicable, – designates available but nonapplicable, ⊕ designates derivable (online computation) and potentially applicable, and an empty cell designates nonavailable or irrelevant.

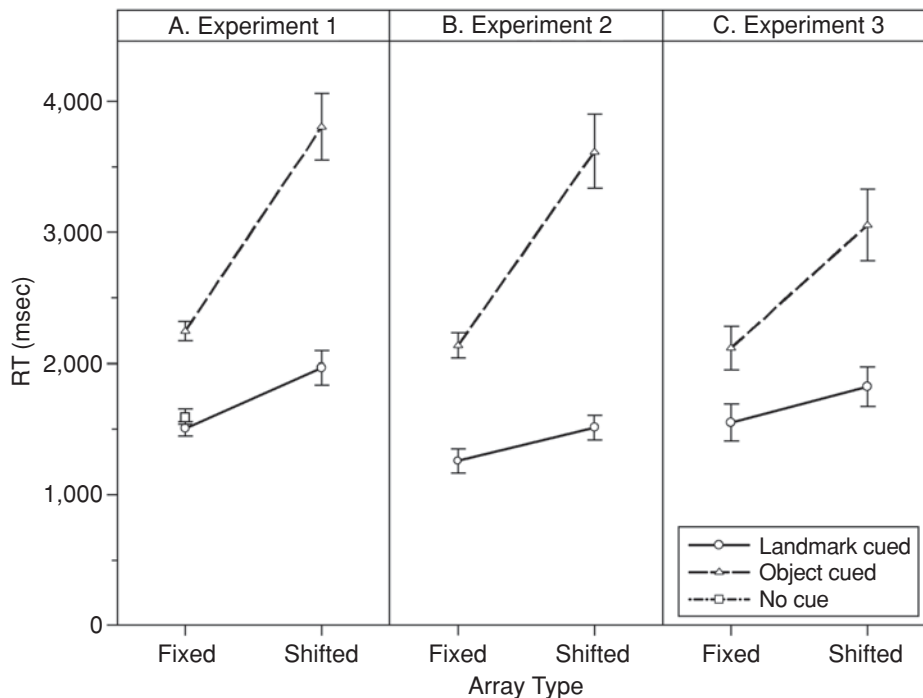


Figure 2. The reaction time (RT) data in the three experiments. The error bars represent standard errors.

pants were able to perform the task in the fixed–no-cue condition with high accuracy (95%). Since, presumably, a screen-based spatial representation had to be used in order to perform the task in this condition and the participants were instructed in the study phase to remember the locations of objects relative to the landmarks, this finding alone suggested that screen-based representations were encoded in the study phase even when the participants were not instructed to do so. These screen-based representations were maintained and later used in the testing phase. This was not surprising, in that the computer screen presented a steady and consistent frame of reference in which object locations could be encoded.

Second, the RT in the fixed–no-cue condition was not significantly different from that in the fixed-landmark condition but was significantly different from that in the shifted-landmark condition. Since the participants were explicitly instructed to encode object locations relative to the landmarks, landmark-based location representations were presumably available in all the testing conditions. In addition, screen-based representations, although available, were apparently invalid and, thus, not applicable in the shifted-landmark condition, due to the shifted array. Therefore, the difference between the fixed–no-cue and the shifted-landmark conditions might, in effect, reflect the difference between the operations supported by screen-based and landmark-based representations.

Third, the significant interaction between cue type (landmark vs. object) and array type (fixed vs. shifted) indicated that the effect of shifting was affected by the cue types. In particular, the RT in the shifted-object con-

dition was significantly longer than that in any other conditions, suggesting that some additional operations occurred in that condition. An examination of the operational differences among conditions shed light on what these operations could be. Specifically, landmark cues (solid squares) were much more perceptually distinct than object cues. In both object-cued conditions, an additional search operation was necessary in order to distinguish the target–distractor pair from the two object retrieval cues. Furthermore, in contrast to the fixed conditions, both shifted conditions required explicit access to spatial relationships, either landmark based or object based. Whereas landmark-based spatial relationships might be directly encoded in the study phase and later directly retrieved in the testing phase, it seems that object-based spatial relationships had to be derived through late computation, because the participants never saw any two objects at the same time during study.

In order to understand these results in a more systematic way, we further differentiated different testing conditions by identifying the available representations underlying each testing condition and examining the different role each representation could play in each condition. This analysis, shown in Table 1, was based on the following three hypotheses. First, the screen-based representations were encoded and available in all five conditions. However, such representations might not always be applicable for later problem solving. In particular, although they were valid and could play a positive role in the fixed array conditions, their availability might actually hurt the performance and result in longer RTs in the shifted array

conditions, since shifting the whole array made these representations invalid for problem solving. Second, whereas landmark-based representations were explicitly encoded in the study phase (because the participants were instructed to do so), object-based representations had to be computed online later, when needed in the testing phase (because no two objects were ever presented at the same time during study). Consequently, object-based representations would be associated with longer RTs. Third, regardless of the array type, landmark-based representations were always valid and applicable in the landmark-cued conditions, and object-based representations were always valid and applicable in the object-cued conditions. As a result, they played a positive role in each of their respective testing conditions.

The analysis shown in Table 1 allows us to systematically examine the effect of each type of representation by comparing different conditions. First, since screen-based representations were the only applicable representations in the fixed-no-cue condition, we could estimate the operational cost associated with this type of representation to be about 1,592 msec. Second, either screen-based or landmark-based representations could be used to solve the fixed-landmark condition. If the two types of representations facilitated each other, we would expect to see shorter RTs in the fixed-landmark condition, as compared with the fixed-no-cue condition. The fact that no such facilitation effect was found suggested two possible retrieval strategies in the fixed-landmark condition: (1) Screen-based representations were dominant and were the only representations used for retrievals, whereas landmark-based representations were simply ignored, and (2) landmark-based representations dominated and were used but had the same time cost as screen-based representations. If we further compare the fixed-landmark and the shifted-landmark conditions, it seems that the first possibility is more plausible. Specifically, note that screen-based representations were available but invalid in the shifted-landmark condition, due to the array shift. If landmark-based representations dominated, we would expect that these nonapplicable screen-based representations had no effect, since both conditions could then be solved by the more dominant landmark-based representations. The finding that there was a significant difference between the two conditions [RT difference = 464 msec; $t(20) = -4.47, p < .001$] suggested that there might be an interference effect from those invalid but dominant screen-based representations on the landmark-based representations.

Furthermore, consider the two object-cued conditions. Again, screen-based representations were available and more dominant than object-based ones. However, they were applicable only in the fixed-object condition, but not in the shifted-object condition. Therefore, we might expect that a similar interference effect among different representations would lead to a significant RT difference between the two conditions. This was what we found [RT difference = 1,558 msec; $t(20) = -7.11, p < .001$]. Finally, let us compare the fixed-landmark and the fixed-object conditions. Since in both conditions screen-based repre-

sentations were available and applicable, we might expect that the two conditions had similar RTs if screen-based representations were solely responsible for the retrievals. However, this was not the case. We found a significant difference of about 741 msec [$t(20) = -10.21, p < .001$]. A careful examination of the difference between the fixed-landmark and the fixed-object conditions indicates that an additional search operation (to distinguish the object cues from the target-distractor pair) was necessary for the object-cued condition, but not for the landmark-cued condition. This was because the landmark cues were more perceptually salient than those object cues. We suspected that the RT difference between the two conditions, 741 msec, was the cost of this search operation.

In sum, the results of Experiment 1 provided reasonable support for our analysis about the interaction among different spatial representations in object location memory. However, they also raised new issues. First, although the results were consistent with our explanation that when the dominant screen-based representations were available but not applicable (invalid), they interfered with less dominant representations and led to longer RTs, we did not yet have a strong test for this explanation. What would happen if we eliminated these dominant screen-based representations? One prediction was that eliminating the dominant and *applicable* representations (e.g., those in the fixed array conditions) would force the participants to resort to the less dominant representations, which would result in longer RTs. The second prediction was that eliminating the dominant but *nonapplicable* representations (e.g., those in the shifted array conditions) would eliminate their interference with the less dominant but applicable representations, which would lead to shorter RTs.

The second issue was the role of search in the object-cued conditions. Since the target-distractor pair and the object cues were all black-and-white drawings, they were not readily distinguishable visually. We hypothesized that a nonspatial visual search component was necessary to identify which drawings were the cues and which ones were the target-distractor pair. In our analysis above, we estimated that the search operation took about 741 msec. Although this explanation was reasonable and convenient, the present experiment did not allow us to specifically test the real effect of this search component. If we eliminated this search component, would we get shorter RTs?

We designed Experiment 2 to explore the first issue and Experiment 3 to explore the second issue.

EXPERIMENT 2

Experiment 2 adopted the same Milner paradigm but differed from Experiment 1 in one major aspect. Specifically, in the study phase of Experiment 1, each object, along with the two landmarks, was always presented at the same screen location during the four-time repetition. This helped explain why screen-based representations were encoded and dominated other representations later in the testing phase. To test this explanation, in Experiment 2, we used a shifted array encoding. That is, in the

study phase, each object was still presented four times, along with the two landmarks, but each time the entire array was shifted to a different screen location. The shifting was done in a fashion similar to that used in the testing phase in the shifted array conditions in Experiment 1. As a result of this manipulation, whereas the object's location with respect to the landmarks remained constant, its absolute screen location kept changing. We expected that this manipulation would effectively destroy the screen-based representations and, consequently, eliminate its interaction with other representations later in the testing phase. Specifically, we predicted that in cases in which screen-based representations would have been applicable, eliminating them would force people to resort to other, less dominant representations, resulting in longer RTs. On the other hand, in cases in which screen-based representations would have played a negative role, eliminating them would reduce the possible interference effect and thus lead to shorter RTs.

Because the fixed-no-cue condition was possible only when screen-based representations were available and adopting a shifted array encoding effectively eliminated screen-based representations, the fixed-no-cue condition was not included in Experiment 2.

Method

Participants. Twelve college students from the University of Texas Health Science Center at Houston, 6 of them female and 6 male, with normal or corrected-to-normal vision, were paid to participate in the experiment.

Apparatus and Materials. Four of the five sets of stimuli used in Experiment 1 were used in this experiment. The stimulus presentation apparatus was identical to that in Experiment 1.

Design and Procedure. Dropping the fixed-no-cue condition resulted in four testing conditions, which were identical to those shown in Figures 1C–1F. Each participant performed in all four conditions in random order. In Experiment 2, the same experimental procedure as that in Experiment 1 was used, except that the array (two landmarks and one object) on each study screen was now shifted, instead of fixed, from trial to trial. The shifting mechanism was identical to that used in the shifted array testing conditions described before.

Note that with a shifted array encoding, the landmarks in the fixed-landmark condition were no longer “fixed” relative to the encoding (i.e., stationed at the same screen locations); they were “fixed” only across all the testing trials. As a result, in both the fixed-landmark and the shifted-landmark conditions, the whole testing array was now essentially shifted relative to the encoding array, and this was how we effectively eliminated the screen-based representations. The same logic applied to the distinction between the fixed-object and the shifted-object conditions.

Results

Accuracy data. The average accuracy for the four testing conditions ranged from 88.8% to 99.5% (see Table 1). There were no differences among the conditions [$\chi^2(3) = 3.11, p > .37$], and there were no significant differences between Experiments 1 and 2 in corresponding conditions [maximal accuracy difference = 5.0%; $\chi^2(1) = 0.07, p > .79$].

RT data. The RT data from all the participants were pooled together and those trials with (1) wrong answers or

(2) RTs that were three standard deviations above the corresponding condition means were designated as invalid (about 9.3% of the total trials) and were eliminated from further analyses.

The valid RT data for all four testing conditions are shown in Figure 2 (see Table 1 for numerical values). A repeated measures ANOVA with a 2 (retrieval cue type: landmark vs. object) \times 2 (array type: fixed vs. shifted) design showed a significant interaction between retrieval cue type and array type [$F(1,11) = 28.77, p < .001$]. The main effects of both cue type and array type were significant [$F(1,11) = 109.77, p < .001$, and $F(1,11) = 37.73, p < .001$, respectively]. Multiple planned paired *t* tests were then conducted to further test the simple effects among different testing conditions. All the tests were again corrected on the basis of Keppel's (1991) modified Bonferroni test and an adjusted significance level of .025 ($3 * .05/6$; see note 2) was adopted. The results showed that the fixed-landmark condition had significantly shorter RTs than did all the other three conditions [minimal RT difference = 253 msec; $t(11) = -4.05, p < .002$]. In addition, the shifted-landmark condition had shorter RTs than did the shifted-object condition [RT difference = 2,108 msec; $t(11) = -9.04, p < .001$], and the fixed-object condition had shorter RTs than did the shifted-object condition [RT difference = 1,484 msec; $t(11) = -5.94, p < .001$]. This general RT pattern among different testing conditions was clearly consistent with the results from Experiment 1.

However, our major interest here was to compare the corresponding four conditions between the two experiments and to see whether eliminating screen-based representations would change the RTs. Four planned comparisons of corresponding conditions in the two experiments were conducted with a Bonferroni-adjusted significance level of .0375 ($3 * .05/4$; see note 2). The results showed that Experiment 2 had shorter RTs in both the fixed-landmark and the shifted-landmark conditions [RT difference = 248 msec, $t(31) = -2.49, p < .02$, and RT difference = 459 msec, $t(31) = -2.51, p < .02$, respectively]. However, there was no significant difference for the fixed-object and the shifted-object conditions between the two experiments (RT differences = 113 and 186 msec, respectively).

Discussion

Experiment 2 was designed to examine the interaction among multiple representations by manipulating their availabilities. (See Table 1 for a representational decomposition for Experiment 2.) We predicted that eliminating screen-based representations would (1) lower the performance (increased RTs) when screen-based representations would have originally been helpful (i.e., fixed array conditions) and (2) improve the performance (reduced RTs) when screen-based representations would have originally been harmful (i.e., shifted array conditions). The results in Experiment 2 partially confirmed our second prediction, but not the first one. Specifically, we found significantly reduced RTs in Experiment 2 in both landmark-cued con-

ditions, regardless of the array type. On the other hand, in both object-cued conditions, the RTs did not change significantly between the two experiments.

This result was surprising, as well as encouraging. On the basis of Table 1, there were four interesting comparisons. First, as we mentioned before, by effectively eliminating the screen-based representations, the fixed-landmark and the shifted-landmark conditions had the identical representational decompositions in Experiment 2. However, the two conditions had a significant RT difference of about 253 msec. It seems that shifting the array around the screen from trial to trial during the testing in the shifted-landmark condition might have cost the participants extra time for landmark localization, which was necessary to effectively use landmark-based representations. Second, let us compare the shifted-landmark conditions in the two experiments. As was predicted, it seems that eliminating the dominant but harmful screen-based representations did help performance (the RT savings was 459 msec). In Experiment 2, the participants were no longer interfered with by those invalid screen-based representations, as in Experiment 1, and could directly resort to the landmark-based representations. Third, a significant time savings of 248 msec in the fixed-landmark conditions from Experiment 1 to Experiment 2 was unexpected. We predicted increased RTs. It appears that eliminating the dominant screen-based representations in Experiment 2 actually helped performance. One possible explanation is that since the participants were explicitly instructed to encode landmark-object relations, these landmark-based representations were strongly represented and could actually support faster responses. In Experiment 1, however, they were not as dominant and convenient as those automatically encoded screen-based representations. When they were released from the dominance of screen-based representations in Experiment 2, they became the only available representations and thus led to faster responses.

Finally, let us compare those object-cued conditions in the two experiments. Although the RT savings from eliminating screen-based representations were 112 and 187 msec for the fixed-object and the shifted-object conditions, respectively, both savings were not significant. Again, we suspected that this had to do with the additional costs associated with search (i.e., distinguishing the target-distractor pair from the object cues) and deriving object-based representations online in those object-cued conditions. As compared with these costs, the effect of eliminating screen-based representations was probably too small, and we lacked adequate statistical power to detect it. We designed Experiment 3 to specifically test the influence of search on performance.

EXPERIMENT 3

Experiment 3 adopted the same Milner fixed array encoding paradigm as that used in Experiment 1 but differed in two aspects (see Figure 3). First, in the object-cued conditions, the two object cues (but not the target-distractor pair) were framed in black-outlined squares. The

purpose of this change was to make the object cues easily distinguishable from the target-distractor pair, so that the effect of the extra search component, we hypothesized, might be reduced. The second difference was to change the landmarks from solid black squares to black-outlined empty squares. This change, along with the first, made the retrieval cues (either landmarks or objects) more visually consistent across different testing conditions.

Method

Participants. Fourteen college students from the University of Texas Health Science Center at Houston, 4 of them female and 10 male, with normal or corrected-to-normal vision, were paid to participate in the experiment.

Apparatus and Materials. Four of the five sets of stimuli used in Experiment 1 were used in this experiment again. The stimulus presentation apparatus was identical to that in Experiment 1.

Design and Procedure. As in Experiment 2, we dropped the fixed-no-cue condition, which resulted in four testing conditions (see Figure 3). Note that the landmarks were now simply frames and that the two object cues in the object-cued conditions were also framed. Each participant performed the tasks in all four testing conditions in random order, and fixed array encoding was adopted in the study phase. Experiment 3 used the same procedure as that in Experiment 1.

Results

Accuracy data. The average accuracy for the four testing conditions ranged from 88.4% to 95.8% (see Table 1). There were no significant differences among the conditions [$\chi^2(3) = 1.29, p > .73$], and there were no significant differences among the three experiments in corresponding conditions [maximal accuracy difference = 5.4%; $\chi^2(2) = 0.65, p > .72$].

RT data. The RT data from all the participants were pooled together, and those trials with (1) wrong answers or (2) RTs that were three standard deviations above the corresponding condition means were designated as invalid (about 11% of the total trials) and were eliminated from further analyses.

The valid RT data for all four testing conditions are shown in Figure 2 (see Table 1 for numerical values). A repeated measure ANOVA with an overall 2 (retrieval cue type: landmark vs. object) \times 2 (array type: fixed vs. shifted) design showed a significant interaction between cue type and array type [$F(1,13) = 8.34, p < .05$]. In addition, the main effects of both cue type and array type were significant [$F(1,13) = 31.00, p < .001$, and $F(1,13) = 22.11, p < .001$, respectively]. Multiple planned paired *t* tests were then conducted to further test the simple effects among those different testing conditions. A Bonferroni-corrected significance level of .025 (3 * .05/6; see note 2) was adopted. The results showed that the fixed-landmark condition had shorter RTs than did the fixed-object and shifted-object conditions [RT difference = 565 msec, $t(13) = -4.16, p < .002$, and RT difference = 1,503 msec, $t(13) = -6.13, p < .001$, respectively]. In addition, the shifted-landmark condition had shorter RTs than did the shifted-object conditions [RT difference = 1,228 msec; $t(13) = -5.02, p < .001$], and the fixed-object condition had shorter RTs than did the shifted-object condition [RT difference = 938 msec;

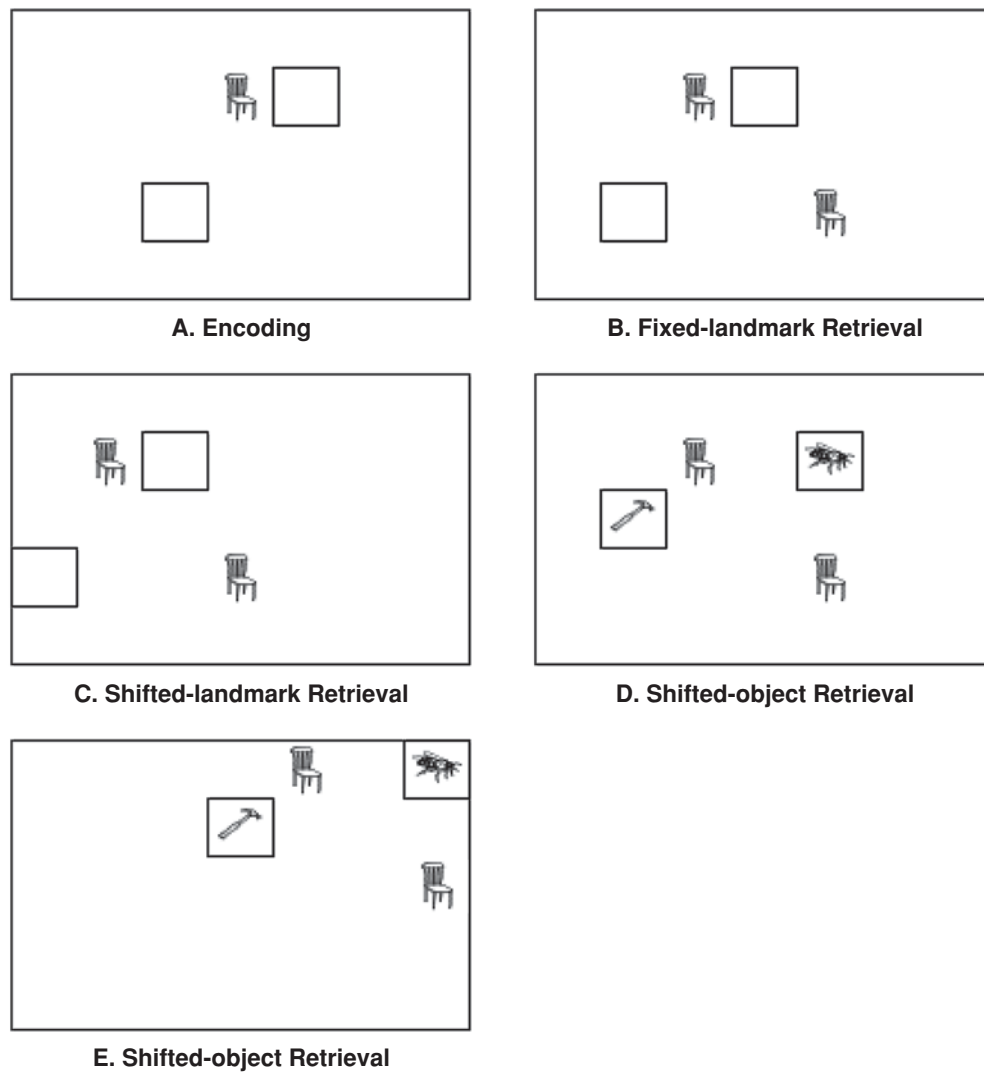


Figure 3. The design of Experiment 3. Panel A represents a study trial, and panels B–E represent various testing conditions. The participants had to identify one chair that was in its original studied location, relative to the landmarks or other studied objects.

$t(13) = -4.62, p < .001$]. Although the difference between the fixed-landmark and the shifted-landmark conditions was not significant [RT difference = 275 msec; $t(13) = -2.03, p < .063$], the general pattern was quite consistent with that in Experiment 1, showing that object-cued conditions took longer than landmark-cued conditions and shifted arrays took longer than fixed arrays.

We further examined the effects of the manipulations we made in Experiment 3 by comparing the results from Experiment 1 and Experiment 3. Four planned comparisons of corresponding conditions in the two experiments were conducted with a Bonferroni-adjusted significance level of .0375 ($3 * .05/4$; see note 2). The results showed no significant differences in the fixed-landmark, shifted-landmark, and fixed-object conditions between the two experiments (RT differences = 46, 143, and 130 msec,

respectively). In addition, the difference in the shifted-object condition did not survive a Bonferroni correction [RT difference = 751 msec; $t(33) = -1.98, p < .056$].

Discussion

A representational decomposition for Experiment 3 was performed, which is shown in Table 1. Since our manipulations in Experiment 3 did not change the underlying representational structures found in Experiment 1, Table 1 shows that the two experiments shared the same representational decompositions. Basically, Experiment 3 replicated the essential results of Experiment 1 and produced interesting new findings. First, note that the only difference between Experiments 1 and 3 in the two basic testing conditions (fixed landmark and shifted landmark) was the change of shapes of the landmarks (from solid

black squares to black-outlined frames). This change did not make a significant difference in performance (RT differences = 46 and 143 msec, respectively), suggesting that the perceptual salience of landmarks might not play a significant role in landmark-cued conditions. This consistency also indicated that the results of the two experiments were quite comparable. Second, when the object-cued conditions between the two experiments are compared, it is clear that the elimination of the search operation (by framing the object cues) did decrease RTs in certain object-cued conditions. Specifically, there was a time savings (751 msec) in the shifted-object condition from framing the object cues. This time savings was remarkably close to our previous estimate of the search cost (741 msec) in Experiment 1.

One unexpected result in Experiment 3 was related to the fixed-object condition. In particular, we expected that eliminating the search operation in this condition would reduce RTs, when compared with the corresponding condition in Experiment 1. Although we did find a time savings of 130 msec, it was not significant. Why did eliminating searching not lead to a larger time savings? One explanation is that the overhead we attributed to the cost of search might actually be a confounded effect of both search and the interaction (interference) between screen-based representations and object-based representations and that the latter was insensitive to the cue-framing manipulation. Interestingly, this explanation was further supported by another observation. When the fixed-landmark and the fixed-object conditions in Experiment 3 (Figures 3B and 3D) are compared, it seems that their only difference was that the frames were empty in Figure 3B but filled with objects in Figure 3D. If screen-based representations indeed dominated in those fixed array conditions, as we have suggested, in both conditions it would have been easy to simply use the dominant screen-based representations to solve the problem, resulting in similar RTs. Apparently, this was not the case. We found that the fixed-object condition took much longer than the fixed-landmark condition [RT difference = 565 msec; $t(13) = 4.16, p < .01$]. If we took this overhead as an estimate of the interaction between screen-based and object-based representations, it now seems that the 741-msec total overhead that we found in Experiment 1 might actually include two parts: about 130 msec for the search cost (the RT difference in the fixed-object conditions between Experiments 1 and 3) and about 565 msec for the interaction cost. In accord with this speculation, previous research has shown that human spatial attentional movement, a central component for search, operates in the range of 100 msec (see Luck, Woodman, & Vogel, 2000, for a review).

GENERAL DISCUSSION

Human spatial thinking is the process of representing and reasoning about spatial relationships among spatial objects and events. Object location memory is an essential aspect of spatial thinking. In the empirical study reported

here, the representational and computational mechanisms underlying human object location memory were explored. Our results show that when multiple spatial representations are simultaneously available for the encoding of object locations, they interact to determine performance.

One of the key features of the Milner paradigm is its simultaneous involvement of multiple spatial representations in performing the object location memory task. By systematically manipulating the testing conditions, the paradigm allows the relative contribution of each representation in the overall spatial performance to be measured and evaluated. We made several extensions of the Milner paradigm with the intention of studying how different object location representations are encoded and interact to solve spatial tasks. The results support the argument that memory for spatial relationships can take multiple forms of representations, each encoded in a different frame of reference. Some of these representations can be acquired in the original encoding phase, such as the screen-based and landmark-based representations. However, many object-based spatial relationships have to be derived later when necessary, resulting in longer RTs in the object-cued conditions. Our results suggest that when multiple forms of representations are simultaneously available, they interact to solve the problem and determine performance. Specifically, those representations that are originally encoded, extensively practiced, and/or conveniently available are dominant. In cases in which they are valid and can be effectively applied, they improve performance and lead to faster responses. On the other hand, in cases in which they are invalid and not directly applicable, people may have to resort to those less dominant representations. Consequently, the availability of these dominant representations may actually interfere with the application of other representations and lead to slower responses.

Although our experimental results support the claim that spatial performance is determined by the interaction among multiple available spatial representations, the exact nature of the interaction clearly needs further investigations. Our experiments seem to suggest that the interaction was more likely unidirectional: Both the facilitation and the interference effects were only from the dominant representations to the less dominant ones. Although this is intuitively appealing, it is important to note that alternative explanations exist. For example, one may argue that the tasks in all of the testing conditions could have been done using screen-based representations, since they are always implicitly encoded during study. To use screen-based locations in the shifted conditions, one could retrieve the original screen-based location of one of the cues and then use the difference between the original and the current locations to compute the angle and distance at which the entire array is shifted. One could then use this information to identify the target by determining which object in the target–distractor pair had undergone the same shift. If our participants had indeed used this approach, we would expect both shifted conditions in Experiment 3 to produce similar RTs, because the two conditions were almost iden-

tical in determining how the array was shifted except for the cues used (landmarks or object cues). However, this was inconsistent with the results of Experiment 3, which showed a significant difference in RT between the two conditions.

We should also consider how the instructions influenced the use of different frames of reference in each testing condition. In our experiments, trials in each testing condition were grouped into sessions, and specific instructions were given for each session. The reason for this was that different testing conditions involved different retrieval cues that the participants had to be informed about. However, to encourage the participants to freely choose whatever representations were available, we made the instructions for each testing condition as general as possible. Specifically, the instructions were of two types. For all fixed testing conditions, the instructions were the same; the participants were instructed to “select the object [of the two identical ones] that is at the same location as you saw in the preceding study session.” No specific clarification about what “the same location” meant was given. For the shifted conditions, the participants were instructed to “select the object [of the two identical ones] whose location relative to the landmarks remains unchanged as in the preceding study session” (for the shifted-landmark condition) or to “select the object [of the two identical ones] whose location relative to the other two objects remains unchanged as in the preceding study session” (for the shifted-landmark condition). We believe that these instructions provided sufficient information for the participants to identify the target object but did not constrain how they were to achieve that and, therefore, served the purpose of the experiments well.

One potential confounding factor in Experiment 1 was the perceptual differences between the landmarks and the objects. Although both the landmarks and the objects were controlled to be 100×100 pixels in size, their differences, in terms of both size and perceptual salience, were apparent. In particular, the landmarks (solid black squares), in general, appeared larger and more salient than the object drawings. Since salience affects both visual search and spatial memory (see Egeth & Yantis, 1997; Itti & Koch, 2001; Theios & Amrhein, 1989; H. Wang et al., 2001), this factor might provide an alternative explanation for the finding that landmark-based conditions were easier than object-based conditions. However, it seems that Experiment 3 provided a test against this hypothesis. In Experiment 3, the landmarks were black frames filled with white, rather than solid black, squares. In addition, every object cue in the object-cued conditions in Experiment 3 was presented inside a 100×100 frame that was exactly the same as a landmark (see Figure 3). As we reported earlier, Experiments 1 and 3 produced comparable results, suggesting that the perceptual differences between the landmarks and the object drawings might not be a critical factor here.

The results from this study fit quite well with the large body of neuropsychological evidence that suggests that

there exist multiple spatial representational systems in the brain. For example, recent evidence suggests that spatial information processing is subserved by multiple separate but interdependent subsystems in the brain, each using distinctive representational schema and, thus, holding different views of space (see H. Wang, et al., 2001 for a review). At least two such systems, the parietal cortex and the hippocampal system, have been identified, with the former specialized for various egocentric representations (such as eye-centered and body-centered spatial representations; e.g., Colby & Goldberg, 1999; Halligan et al., 2003) and the latter for allocentric representations (such as place cells, cognitive maps, and episodic memory; see, e.g., Redish, 1999). Although previous research has typically emphasized their differences and independence, the present study suggests that they may work closely together and interact to solve spatial problems (see also Burgess, Jeffery, & O’Keefe, 1999b).

In sum, the three experiments reported in this article examined the representational nature of human object location memory. The results show that multiple frames of reference can be used to encode spatial relationships among objects and that, when they are simultaneously available, they interact to determine performance. However, we recognize that our experiments examined only simple two-dimensional computer-screen-based “toy” spaces, which are quite different from the real three-dimensional and navigable environment. Although our results are, in general, consistent with previous research with the real-world environment (e.g., McNamara, 2003), further efforts are necessary to extend our results to explain real-world performance.

REFERENCES

- BREWER, B., & PEARS, J. (1993). Frames of reference. In R. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 25-30). Oxford: Blackwell.
- BURGESS, N., JEFFERY, K. J., & O’KEEFE, J. (Eds.) (1999a). *The hippocampal and parietal foundations of spatial cognition*. New York: Oxford University Press.
- BURGESS, N., JEFFERY, K. J., & O’KEEFE, J. (1999b). 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). New York: Oxford University Press.
- COLBY, C. L., & GOLDBERG, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, **22**, 319-349.
- DIWADKAR, V. A., & MCNAMARA, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, **8**, 302-305.
- DRIVER, J., & POUGET, A. (2000). Object-centered visual neglect, or relative egocentric neglect? *Journal of Cognitive Neuroscience*, **12**, 542-545.
- EASTON, R. D., & SHOLL, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 483-500.
- EGETH, H. E., & YANTIS, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, **48**, 269-297.
- FRANKLIN, N., TVERSKY, B., & COON, V. (1992). Switching points of view in spatial mental models. *Memory & Cognition*, **20**, 507-518.
- GALLISTEL, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- GARNHAM, A. (1989). A unified theory of the meaning of some spatial relational terms. *Cognition*, **31**, 45-60.

- HALLIGAN, P. W., FINK, G. R., MARSHALL, J. C., & VALLAR, G. (2003). Spatial cognition: Evidence from visual neglect. *Trends in Cognitive Sciences*, *7*, 125-133.
- ITTI, L., & KOCH, C. (2001). Computational modeling of visual attention. *Nature Reviews*, *2*, 194-203.
- JOHNSRUDE, I., OWEN, A. M., CRANE, J., MILNER, B., & EVANS, A. C. (1999). A cognitive activation study of memory for spatial relationships. *Neuropsychologia*, *37*, 829-841.
- KEPPEL, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- KING, J. A., BURGESS, N., HARTLEY, T., VARGHA-KHADEM, F., & O'KEEFE, J. (2002). Human hippocampus and viewpoint dependence in spatial memory. *Hippocampus*, *12*, 811-822.
- KLATZKY, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge* (pp. 1-17). New York: Springer.
- LANSDALE, M. W. (1998). Modeling memory for absolute location. *Psychological Review*, *105*, 351-378.
- LEVINSON, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 109-169). Cambridge, MA: MIT Press.
- LOGAN, G. D., & SADLER, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 493-529). Cambridge, MA: MIT Press.
- LUCK, S. J., WOODMAN, G. F., & VOGEL, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*, 432-440.
- 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: Springer.
- MILNER, B., JOHNSRUDE, I., & CRANE, J. (1997). Right medial temporal lobe contribution to object-location memory. *Philosophical Transactions of the Royal Society of London: Series B*, *352*, 1469-1474.
- MOU, W., & MCNAMARA, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 162-170.
- NAVEH-BENJAMIN, M. (1987). Coding of spatial location information: An automatic process? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 595-605.
- NAVEH-BENJAMIN, M. (1988). Recognition memory of spatial location information: Another failure to support automaticity. *Memory & Cognition*, *16*, 437-445.
- NEWCOMBE, N. S. (2002). Spatial cognition. In C. R. Gallistel (Ed.), *Stevens' Handbook of experimental psychology: Vol. 2. Memory and cognitive processes* (3rd ed., pp. 113-163). New York: Wiley.
- NEWCOMBE, N. S., & HUTTENLOCHER, J. (2000). *Making space: The development of spatial representation and reasoning*. Cambridge, MA: MIT Press.
- O'KEEFE, J. (1993). Kant and the sea-horse. In N. Eilan, B. Brewer, & R. McCarthy (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 43-64). Oxford: Blackwell.
- O'KEEFE, J. (1996). The spatial prepositions in English, vector grammar, and the cognitive map theory. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 277-316). Cambridge, MA: MIT Press.
- O'KEEFE, J., & NADEL, L. (1978). *The hippocampus as a cognitive map*. New York: Oxford University Press.
- OWEN, A. M., MILNER, B., PETRIDES, M., & EVANS, A. C. (1996). A specific role for the right parahippocampal gyrus in the retrieval of object-location: A positron emission tomography study. *Journal of Cognitive Neuroscience*, *8*, 588-602.
- PALMER, B. (2003). Linguistic frame of reference reconsidered. In P. Collins & M. Amberber (Eds.), *Proceedings of the 2002 Conference of the Australian Linguistics Society*. Available at www.als.asn.au.
- REDISH, A. D. (1999). *Beyond the cognitive map: From place cells to episodic memory*. Cambridge, MA: MIT Press.
- SHELTON, A. L., & MCNAMARA, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review*, *4*, 102-106.
- SHOLL, M. J., & BARTELS, G. P. (2002). The role of self to object updating in orientation-free performance on spatial-memory tasks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 422-436.
- SHOLL, M. J., & NOLIN, T. L. (1997). Orientation specificity in representations of place. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 1494-1507.
- SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 174-215.
- SUN, Y., WANG, H., & JOHNSON, T. R. (2004). Spatial updating in intrinsic frames of reference. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society* (pp. 1285-1290). Mahwah, NJ: Erlbaum.
- TALMY, L. (1983). How languages structure space. In H. Pick & L. Acredolo (Eds.), *Spatial orientation: Theory, research and application* (pp. 225-282). New York: Plenum.
- THEIOS, J., & AMRHEIN, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review*, *96*, 5-24.
- TOLMAN, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189-208.
- TVERSKY, B. (1996). Spatial perspective in descriptions. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 463-492). Cambridge, MA: MIT Press.
- TVERSKY, B. (2000). Remembering spaces. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 363-378). New York: Oxford University Press.
- VALIQUETTE, C. M., MCNAMARA, T. P., & SMITH, K. (2003). Locomotion, incidental learning, and the selection of spatial reference systems. *Memory & Cognition*, *31*, 479-489.
- WALLER, D., MONTELOB, D. R., RICHARDSON, A. E., & HEGARTYA, M. (2002). Orientation specificity and spatial updating of memories for layouts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 1051-1063.
- WANG, H., JOHNSON, T. R., & ZHANG, J. (2001). *The mind's views of space*. In L. Chen & Y. Zuo (Eds.), *Proceedings of the Fourth International Conference of Cognitive Science* (pp. 191-198). Beijing: University of Science & Technology of China Press.
- WANG, R. F., & SIMONS, D. J. (1999). Active and passive scene recognition across views. *Cognition*, *70*, 191-210.
- WANG, R. F., & SPELKE, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, *77*, 215-250.
- WANG, R. F., & SPELKE, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences*, *6*, 376-382.

NOTES

1. Note that in our context, landmarks differ from objects (drawings) in two aspects. First, landmarks, as solid black squares, are perceptually distinctive from namable objects. Second and more important, in the study phase, landmarks always co-appear with each object on each encoding screen (in contrast, objects never co-appear with other objects in this phase). These features make the landmarks visual and spatial anchors. In principle, however, any entity on the screen can serve as a landmark as long as it is distinguishable and constantly available, so that other entities' locations can be easily encoded with respect to it.

2. According to Keppel (1991, pp. 169-170), a Bonferroni-corrected significance level can be determined by $(df * \alpha) / c$, where df is the number of orthogonal comparisons, c is the number of planned comparisons, and α is the standard significance level. In this case, we have five conditions to compare. So $df = 5 - 1 = 4$, and $c = 5 * (5 - 1) / 2 = 10$. With $\alpha = .05$, we have the Bonferroni-corrected significance level = $(4 * .05) / 10 = .02$.

APPENDIX

The drawings used for stimuli were selected from Snodgrass and Vanderwart (1980). They are listed in Figure A1. There is no significant difference among different groups of stimuli in several major semantic characteristics, such as name agreement, image agreement, familiarity, and frequency (Snodgrass & Vanderwart, 1980).










| Set 1 | Set 2 | Set 3 | Set 4 | Set 5 |
|---|---|---|---|--|
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Figure A1. The five sets of object drawings used as stimuli in all the experiments, selected from Snodgrass and Vanderwart (1980).

(Manuscript received April 4, 2003;
revision accepted for publication November 14, 2004.)