A Study of Object-Location Memory

Hongbin Wang (Hongbin.Wang@uth.tmc.edu)
Todd R. Johnson (Todd.R.Johnson@uth.tmc.edu)
Jiajie Zhang (Jiajie.Zhang@uth.tmc.edu)
Yue Wang (Yue.Wang@uth.tmc.edu)
School of Health Information Sciences, University of Texas Health Science Center at Houston
7000 Fannin, Suite 600, Houston, TX 77030 USA

Abstract

This paper aims to study the representational nature of human object-location memory. Two experiments are reported, including both performance data and eye movement data. The results show that multiple allocentric frames of reference are used to encode spatial relationships among objects and late computation in object-location memory retrieval in object-cued conditions is often inevitable. The implications on developing a general model of human spatial cognition are discussed.

Introduction

One important aspect of human spatial memory has to do with remembering the location of objects relative to each other. For example, you might recall that the book you read last night is on your office desk between your computer and the desk lamp. This type of memory for spatial relationships is an essential component of a more general type of memory for spatial layout and is obviously critical for many spatial tasks including locating and navigation (see Tversky, 2000, for a review).

It is not clear, however, how spatial relationships among objects are encoded in memory. While it seems apparent that allocentric frames of reference (i.e., locations defined relative to external objects) rather than egocentric frames of reference (i.e., locations relative to the observer self) are often used to describe these spatial relationships, the representational and computational nature of this description is controversial (see Hunt & Waller, 1999; Klatzky, 1998). For example, are object-based spatial relationships encoded and stored directly (early computation)? Or do they have to be inferred much later at the retrieval stage (late computation)? What factors determine which representational scheme is used?

In this paper we report two experiments we conducted to directly address these issues. The results show that multiple allocentric frames of reference are used to encode spatial relationships among objects and late computation in object-location memory retrieval in object-cued conditions is often inevitable.

This paper consists of three major parts. In the first section, the experimental paradigm is briefly introduced. In the second section, the experiments are reported, including both performance data and eye movement data. In the final section, we briefly discuss our ongoing work of developing a computational model of the object-location memory and its implications on modeling human spatial cognition in general.

The Experimental Paradigm

We adopted an experimental paradigm developed by Milner and colleagues in the 1990s, which we call the Milner paradigm (e.g., Johnsrude, Owen, Crane, Milner, & Evans, 1999; Milner, Johnsrude, & Crane, 1997; Owen, Milner, Petrides, & Evans, 1996). Though their focus was on neuroimaging studies of the brain foundations of object-location memory, the Milner paradigm offers an elegant experimental design that allows a systematic evaluation of multiple schema for representing spatial relationships. In addition, the availability of neuroimaging data provides invaluable constraints on both understanding behavioral results and developing computational models (e.g., Wang, Johnson, & Zhang, 2001).

There are two phases in the Milner paradigm. In the encoding phase (Figure 1A), eight drawings (objects) are individually presented on a computer screen to subjects, with each drawing accompanied by two landmarks (solid squares). Subjects are asked to remember the locations of drawings, relative to the landmarks. In the retrieval phase, subjects are presented some cues plus two identical drawings. One of the two identical drawings (target) is presented in its original location, and the other one (noise) is presented in a different location (or more accurately, it occupies the original location of another object). Subjects are required to perform a forced-choice recognition of the target, relative to the cues. Milner and colleagues originally used four retrieval conditions:

1. In the fixed-landmark condition (Figure 1C), the two landmarks were presented as cues, along with the target-noise pair. The absolute location of landmarks and objects on the screen was unchanged from their original encoding positions.
2. In the shifted-landmark condition (Figure 1D), the two landmarks were presented as cues, along with the target-noise pair. Though the spatial relationships among the landmarks/drawings remained unchanged, the absolute locations of the landmarks/drawings on the screen were shifted.
3. In the fixed-object condition (Figure 1E), two encoded drawings instead the two landmarks were

presented as cues, along with the target-noise pair. The absolute locations of drawings on the screen were unchanged.

4. In the shifted-object condition (Figure 1F), two encoded drawings instead the two landmarks were presented as cues, along with the target-noise pair. Though the spatial relationships among the drawings remained unchanged, the absolute locations of the drawings on the screen were shifted.

One significant feature of the Milner paradigm is that it simultaneously involves multiple spatial representations, including screen-based, landmark-based, and object-based. While landmark-based representations are perceptually accessible in the encoding phase (because an object was always presented along with the two landmarks in the encoding phase), object-based representations are not (because no two objects are presented at the same time in the encoding phase). Therefore, this fact alone might suggest that object-based retrieval would be harder than landmark-based retrieval. Systematic alignment of the different testing conditions allowed Milner and colleagues to use a subtraction method to determine the brain areas that dominate in the different test conditions.

Behavioral data was only briefly reported in Johnsrude et al (1999). It was found that the shifted-object condition was harder (e.g., longer RTs and lower accuracy) than any other conditions, which did not differ from each other. Neuroimaging data suggested that object-location memory in general involved the parahipocampal system, and the shifted conditions, as compared to the respective fixed-conditions, activated the posterior inferotemporal cortex. Both areas have been believed to subserve important functions of spatial cognition (e.g., Burgess, Jeffery, & O'Keefe, 1999).

Experiment 1

In Experiment 1 we added one more testing condition to the original Milner paradigm. In this additional condition, called the fixed-nocue condition, no cues were presented along with the target-noise pair in the retrieval phase. Subjects had to make the forced-choice based solely on the absolute location of objects on the screen. This condition was added to explicitly test the effect of screen-based spatial representations in location retrieval.

Another purpose of experiment 1 was to collect eye movement data. Both perceptually encoding and cognitively computing spatial relationships invite eye movements. The trace of natural eye movements during the task provides an indication of the deployment of attention (e.g., Corbetta et al., 1998) and may shed light on the underlying spatial representations and operations (e.g., Colby & Goldberg, 1999).

Subjects, Apparatus, and Materials

21 subjects, 8 females and 13 males, with normal or corrected-to-normal vision, were paid to participate in the experiment. Five sets of stimuli (each consisting of eight drawings) were created using digitized black and white representational drawings of common objects, selected from the database of (Snodgrass & Vanderwart, 1980). The drawings, 100x100 pixels in size, were presented against a white background on a 19'' VGA monitor with a resolution of 1024x768. The monitor was in front of the subjects within 2 feet. Subjects were asked to respond by clicking with a mouse which was within comfortable reach. 11 subjects wore a head-mounted ISCAN eye-tracker while they were doing the experiment.

Design and Procedure

Each subject performed all 5 experimental conditions, each with a different stimulus set. The design is illustrated in Figure 1.

In each encoding trial, subjects were presented one drawing and two landmarks and were instructed to remember the location of the drawing relative to the landmarks. Subjects clicked the drawing to go on to the next trial. There were 32 encoding trials in each condition, with each drawing presented 4 times. The presentation order was randomized. During the study subjects did not know which testing condition would follow.

In each retrieval trial, subjects were presented the cues and the target-noise pair according to the testing condition. Subjects were instructed to choose the target, by clicking, as quickly as possible and as accurately as possible. As soon as the subjects clicked, the next trial was presented. Each drawing was tested 4 times.

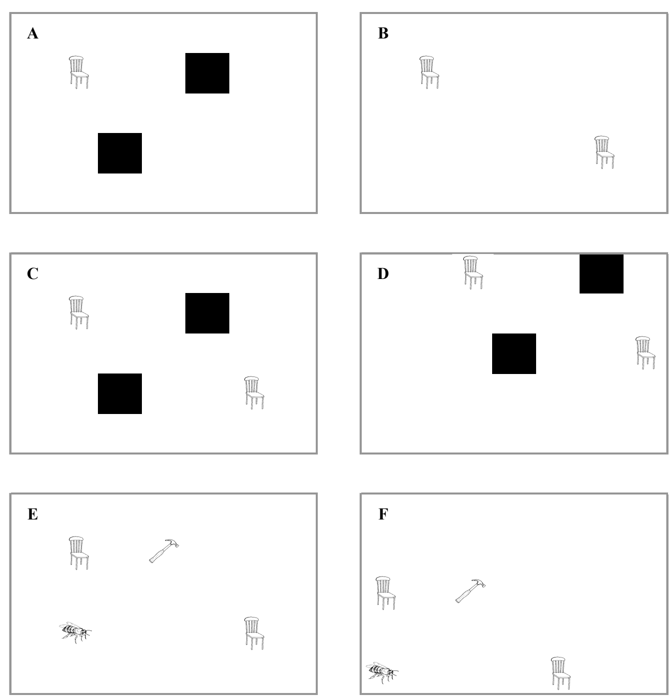

Figure 1. The design of Experiment 1. A, an encoding trial; B, fixed-nocue retrieval; C, fixed-landmark retrieval; D, shifted-landmark retrieval; E, fixed-object retrieval; F, shifted-object retrieval.

Results

Accuracy data. The average accuracy was at least 93%, and there was no difference among the 5 testing conditions.

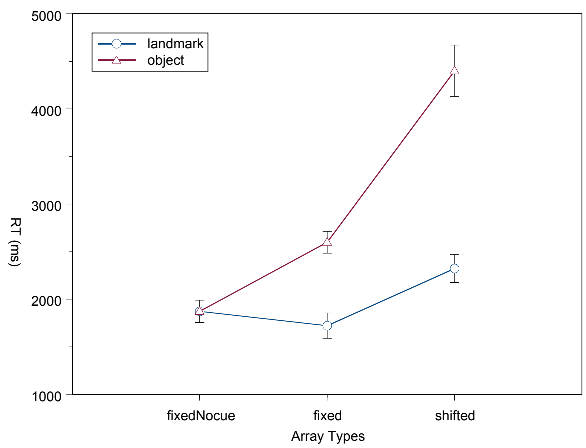

Figure 2. RT data in Experiment 1. The error bars represent 95% confidence intervals.

RT data. The reaction time data is shown in Figure 2. An ANOVA shows a significant interaction between the cue type (landmark vs object) and the array time (fixed vs shifted). In addition, a post-hoc comparison shows that the shifted-object condition takes significantly longer than any other conditions, consistent with Johnsrude et al (1999) results.

Eye movement data. Eye movements are needed to search the scene and measure spatial relationships. The number of eye fixations in each trial is counted and reported here. The result is shown in Figure 3. It is interesting to note that the eye fixation pattern is remarkably similar to the RT pattern, indicating that the number of eye fixations is a good predictor of RT.

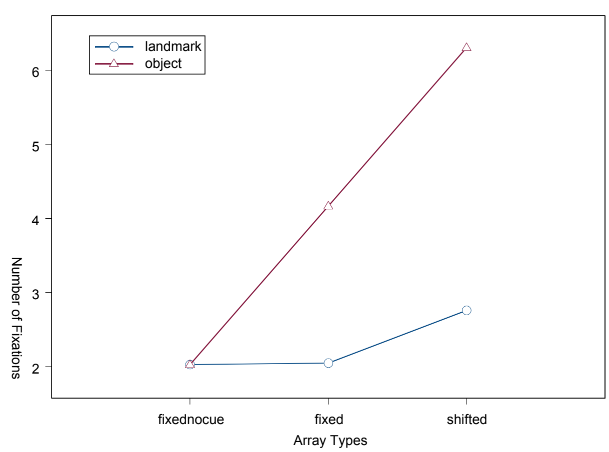

Figure 3. The number of eye fixations in Experiment 1.

Summary & Discussion

Experiment 1 resulted in two major findings that have not been reported by Milner and colleagues. First, the reaction time in the fixed-nocue condition was not different from that in the fixed-landmark condition. Since a screen-based spatial representation had to be used in order to perform the fixed-nocue condition, this result indicates that a screen-based spatial representation might be *implicitly* encoded and stored (because subjects were specifically instructed to pay attention to the drawing's location relative to the landmarks), and be adopted to perform the fixed-landmark condition. This hypothesis was further supported by the eye movement data. The number of eye fixations in both conditions was about 2, the minimal fixations needed to identify the target if a conservative check-both-target-and-noise-before-click strategy was used. The eye movement traces also indicated that subjects often ignored landmarks in the fixed-landmark condition.

Second, the significant interaction between the cue type (landmark vs object) and the array type (fixed vs shifted) was surprising. The reaction time in the shifted-object condition was significantly longer than that in any other conditions (the RT in the shifted-object condition was about 1800ms, 2100ms, and 2700ms longer than that in the fixed-object, shifted-landmark, and fixed-landmark conditions, respectively, see also Table 1), indicating some additional operations occurred in that condition. An analysis of the computational differences among conditions sheds light on what these operations could be. a) 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-noise pair from the two object cues. b) Compared to the fixed conditions, both shifted conditions required explicit access of spatial relationships, either landmark-based or object-based. While landmark-based spatial relationships might be directly encoded in the encoding phase and later directly retrieved in the retrieval phase, it seems that object-based spatial relationships had to be derived through late computation because subjects never saw any two objects at the same time.

Eye movement data, however, indicated that this hypothesis might be oversimplified. In the encoding phase, we quite often observed that subjects moved his/her eyes back and forth between the currently presented object and the location of the previously displayed object (in the previous trial, which has already disappeared), indicating some form of object-based spatial relationships might be encoded directly and quite early. In general, however, it seems likely that shifted-object conditions involved quite extensive late computation in determining object-based spatial relationships.

We speculate that a race model can be used to explain the data. Specifically, when multiple types of representations for spatial relationships are available to solve the task at hand, they compete. Though often the representation that affords easiest operations dominates, sometimes they interfere with each other. A decomposition of the representations/operations for each condition is summarized in Table 1. It seems that the race model explains the RT data reasonably well.

Table 1: A representational decomposition

		Accessible representations/operations		
	RT (ms)	Early computation	Late computation	Addn. ops
Fixed-nocue	1874	Screen-based		
Fixed-landmark	1723	Screen-based Landmark-based		
Fixed-object	2599	Screen-based	Object-based	Search
Shifted-landmark	2324	Landmark-based		
Shifted-object	4402		Object-based	Search

Experiment 1 raised two issues. The first one is the role of search in the object-cued conditions. Since the target-noise pair and the object cues are visually indistinguishable, a non-spatial visual search component is necessary to perform the task. The search component was a free parameter in the above race model that could be estimated but it obviously confounded the results. It would be useful to eliminate this confound. The second issue also has to do with the object-cued conditions. In Experiment 1, the two objects that were chosen to be cues in each trial were randomly selected from all possible objects (i.e., those that were not the target-noise pair). This made the task hard in the sense that all possible object-based spatial relationships (there were 7^8 of them!) might be relevant in the retrieval phase. This was in sharp contrast with the landmark-cued conditions, which had only 16 relevant spatial relationships (8 for each landmark). Therefore, it might be the pure number of relevant spatial relationships but not the late computation of object-based representations that made the object-cued conditions more difficult. We designed experiment 2 to explore these two issues.

Experiment 2

Experiment 2 adopted the same Milner paradigm, but differed from Experiment 1 in three aspects. First, the landmarks were changed from solid black squares to a white-filled black square. Second, in the object-cued conditions, the object cues were framed in a black squared to make them visually salient. The purpose of the change was to eliminate the search component in retrieval. Third, we added two more object-cued conditions, which we called consistent mapping conditions. In these conditions, instead of selecting two cue objects at random for every trial, the two objects were selected at the beginning of the testing session and consistently served as the object cues for every trial in that session. This change greatly reduced the relevant spatial relationships and could be viewed as a middle condition between the pure object-cued condition and the pure landmark-based condition.

These changes resulted in six testing conditions, as shown in Figure 4. 14 subjects were paid to participate in the experiment.

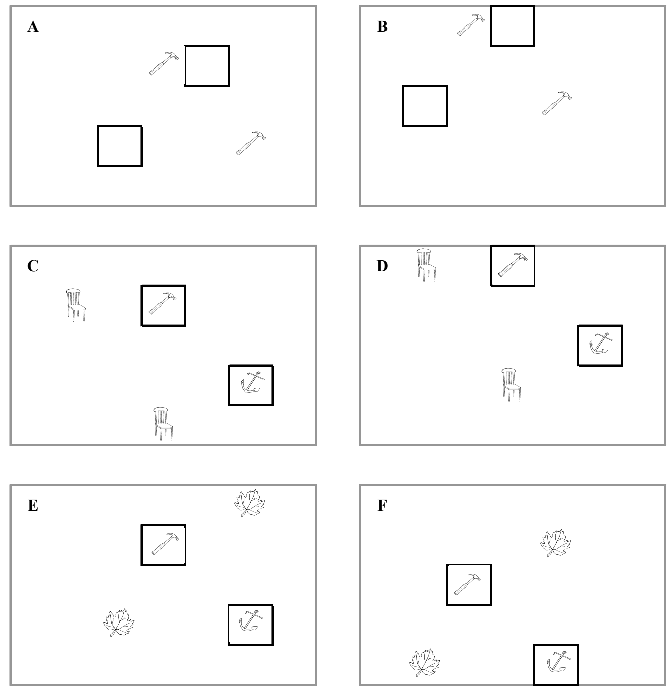

Figure 4. The design of Experiment 2. A, fixed-landmark retrieval; B, shifted-landmark retrieval; C, fixed-object retrieval; D, shifted-object retrieval; E, fixed-object consistent mapping retrieval; F, shifted-object consistent mapping retrieval.

Results & Discussion

Accuracy data. The average accuracy was at least 88%, and there was no difference among the 6 testing conditions.

RT data. The reaction time data is shown in Figure 5. An ANOVA reveals similar effects to Experiment 1, including the significant interaction between cue types (object vs landmark) and array types (fixed vs shifted).

The effects of the two manipulations we adopted in Experiment 2 were evident. First, combining the results from both experiments, it is clear that the elimination of the search operations (by framing the object cues) did decrease reaction time in certain object-cued conditions. However, this time saving had a surprising interaction with the array types. Specifically, while the time saving in the fixed-object condition was not significant (2599ms in Experiment 1 vs 2494ms in Experiment 2) the saving was significant in the shifted-object condition (4402ms in Experiment 1 vs 3548ms in Experiment 2). It is not so obvious how to explain this interaction. Second, the manipulation of consistent mapping in object-cued conditions also reduced the reaction time. However, again, a reliable interaction with array types was found. While the time reduction was about 450ms in the fixed-object conditions (2494ms vs 2044ms), the reduction was about 900ms in the shifted-object conditions (3548ms vs 2640ms). Similarly, it is not obvious how this interaction occurred without a detailed computational model.

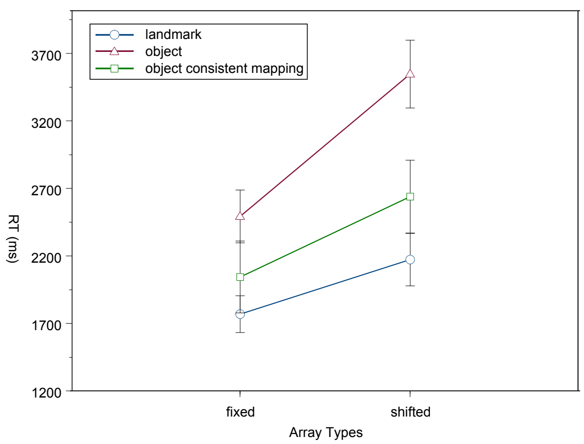

Figure 5. RT data (in ms) in Experiment 2. The error bars represent 95% confidence intervals.

Eye movement data. Similar to Experiment 1, eye movement data corresponded quite well with the reaction time data (see Figure 6). Fewer numbers of eye fixations were observed in the fixed array conditions than in the shifted array conditions. In addition, the object-cued conditions induced more eye fixations than the landmark-cued conditions. In particular, both the elimination of the search component and consistent mapping in object-cued conditions significantly reduced the number of eye fixations (by about 1 and 1.5, respectively), indicating both manipulations successfully reduced the efforts of object recognition and late computation of spatial relationships.

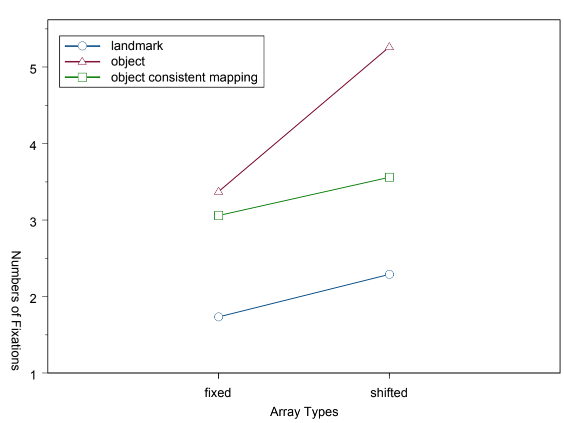

Figure 6. The number of eye fixations in Experiment 2.

General Discussion

Memory for object-location is an essential aspect of human spatial memory. However, the underlying representational mechanisms and computational operations are controversial. The empirical study reported here adopted and extended the Milner paradigm and produced interesting data toward a better understanding of the problem. In this section, we would like to discuss three issues related to the implications of the study and the future work.

First, the current study supports 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 may result from an early computation, often due to a direct perceptual experience in the early encoding phase. These representations can be encoded implicitly, such as the screen-based representations, or explicitly, such as the landmark-based representations and some object-based representations (e.g., spatial relationships between objects presented in consecutive trials). However, most of the object-based spatial relationships have to be inferred when necessary through a late computation, resulting in longer reaction time in the object-cued conditions. When multiple forms of representations are simultaneously available, a race model seems plausible. The processes supported by each representation compete with each other, and typically the one that affords fast response dominates. Eye movement results support this hypothesis

Second, the results from the current study are also consistent with the neuropsychological evidence that suggests there exist multiple spatial representational systems in the brain (e.g., Burgess et al., 1999; Wang et al., 2001). The PET imaging results from Milner and colleagues (1997) revealed that when object-location memory is retrieved, brain activity increases in the parahippocampal system, an area that is generally believed to subserve allocentric spatial representations.

Finally, while the current study generated interesting results, it is clear that to fully understand these results a detailed computational model is necessary. Questions about how multiple forms of spatial relationships are represented and how they interact can be better explored only when a computational model is developed. Efforts are being taken to develop such a model in the Act-R cognitive architecture (Anderson & Lebiere, 1998). The long-term goal is to develop a framework that can be used to model human spatial cognition in general, including object-location memory and spatial layout memory.

Acknowledgments

This work is supported by a grant from the Office of Naval Research (Grant No. N00014-01-1-0074).

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Press.

Burgess, N., Jeffery, K. J., & O'Keefe, J. (Eds.). (1999). *The hippocamal and parietal foundations of spatial cognition*. 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.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Petersen, S. E., Raichle, M. E., Essen, D. C. V., & Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron, 21*, 761-773.

Hunt, E., & Waller, D. (1999). *Orientation and wayfinding: A review.*Unpublished manuscript, Arlington, VA.

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.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In K. F. Wender (Ed.), *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge*. New York: Springer-Verlag.

Milner, B., Johnsrude, I., & Crane, J. (1997). Right medial temporal-lobe contribution to object-location memory. *Phil. Trans. R. Soc. Lond. B, 352*, 1469-1474.

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.

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.

Tversky, B. (2000). Remembering spaces. In F. I. M. Craik (Ed.), *The Oxford handbook of memory*. New York: Oxford University Press.

Wang, H., Johnson, T. R., & Zhang, J. (2001). *The mind's views of space*. Paper presented at the Fourth International Conference of Cognitive Science.