

Michel-Ange Amorim  
amorim@ccr.jussieu.fr  
Collège de France  
Paris, France

Ben Trumbore  
Cornell University  
Ithaca, New York

Pema L. Chogyen  
Columbia University  
New York

# Cognitive Repositioning inside a Desktop VE: The Constraints Introduced by First- versus Third-Person Imagery and Mental Representation Richness

---

## Abstract

Cognitive repositioning is crucial for anticipating the content of the visual scene from new vantage points in virtual environments (VE). This repositioning may be performed using either a first- (immersive-like) or a third-person imagery perspective (via an imaginary avatar). A three-phase study examined the effect of mental representation richness and imagery perspective on the anticipation of new vantage points and their associated objects inside an unfamiliar but meaningfully organized VE. Results showed that the initial level of encoding affects the construction of spatial knowledge, whose exploration is then constrained mostly by the imagery perspective that has been adopted, and the spatial arrangement of the environment. A third-person perspective involves mental extrapolation of directions with the help of a scanning process whose rate of processing is faster than the process used to generate the missing 3-D representation of first-person perspectives. Finally, anticipation of a new vantage point precedes access to its associated object mainly when adopting a first-person perspective for exploring the environment. These findings may prove to be of potential interest when defining cognitively valid rules for real-time automatic camera control in VEs.

## 1 Introduction

New 3-D tools are increasing the use of shared VEs for the purposes of entertainment, learning, and collaborative work (Roehl, 1998). These environments are populated with avatars (Paniaras, 1997), raising challenging issues for real-time technology, cognitive psychology, and social science. This study is concerned with problems related to spatial orientation and perspective change—in other words, “cognitive repositioning” (Young, 1989), a crucial process for the anticipation of VE aspect from new vantage points.

This study was designed to address three issues. First, we examined the effect of encoding levels on the construction of a spatial knowledge issued from an unfamiliar virtual environment. Second, we evaluated the constraints introduced by the type of imagery perspective used to access this spatial knowledge,

---

1. Requests for reprints should be sent to Michel-Ange Amorim, Laboratoire de Physiologie de la Perception et de l'Action, Collège de France, CNRS, 11 place Marcelin Berthelot, 75005 Paris, France.

that is, either a first- (immersive-like) or a third-person imagery (via an imaginary avatar). Finally, we tested the possible effect of both representation richness and imagery perspective on the anticipation of new vantage points or their associated objects in an unfamiliar VE.

Each of these issues will be presented in more detail in the next paragraphs on the basis of the psychological literature and followed by an overview of the study.

### 1.1 First- versus Third-person Mental Imagery

Interestingly, shortly after Piaget and Inhelder's (1956) pioneering work in children, perspective-taking referred to the "ability to imagine or represent how objects look relative to one another from *another person's point of view*" [the author's emphasis] (Cox, 1977). This "person" would usually be a doll (Fishbein, Lewis, & Keiffer, 1972; Light & Nix, 1983) and sometimes even an animal, such as a horse (Huttenlocher & Presson, 1973). In other words, at that time, perspective change would involve what sport psychologists (Mahoney & Avenier, 1977) have termed "third-person" imagery: a perspective that refers to "seeing oneself" from the outside as an external observer standing back with respect to the visual scene with a larger field of view available onto the environment. We have to wait for Hintzman, O'Dell, and Arndt (1981) to find paradigms in which subjects are supposedly "immersed" inside a configuration of objects (Rieser, 1989; Young, 1989; Presson & Montello, 1994; Easton & Sholl, 1995; May & Wartenberg, 1995) and have to imagine their own perspective change using a "first-person" imagery. Usually, the to-be-imagined perspective change would be that of a self-rotation in place of the observer.

Although in both cases the final answer is one of an observer's perspective on a visual scene, what distinguishes both types of task are the mental imagery processes leading to the answer. In third-person imagery, the subject is engaged in more of a "gaze tour" (Ehrich & Koster, 1983) and extrapolates the content of the visual scene faced by the imaginary avatar through attention shifting onto the VE. Otherwise, in first-person imagery, the subject has to imagine the changing visual

scene with his displacement and must reconstruct the spatial configuration missing in the visual scene (Kosslyn, 1980).

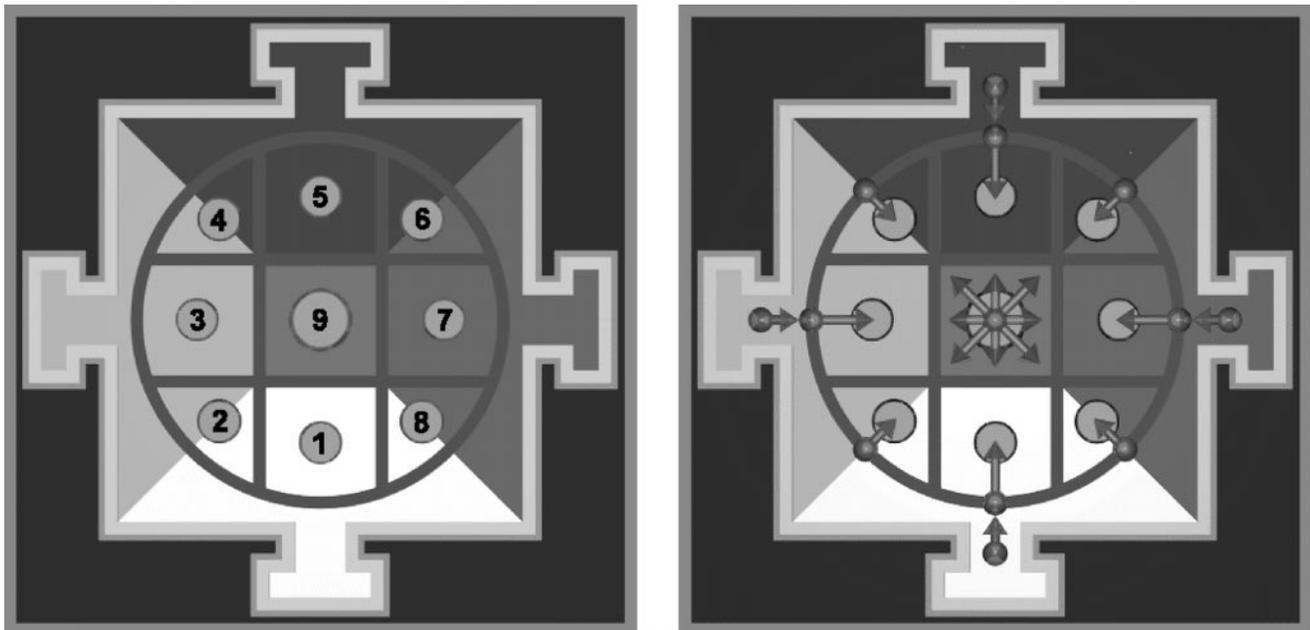
First- and third-person perspectives are popularly illustrated in 3-D games (Bauman, 1996). However, little is known about the effect of these perspectives on cognitive repositioning inside an unfamiliar VE. Therefore, the study may prove to be of special interest for research on real-time automatic camera control in VEs (Drucker & Zeltzer, 1995; He, Cohen, & Salesin, 1996; Bares, Grégoire, & Lester, 1998) which is based on a variety of assumptions about the cognitive skills of a VE user. Also, it may constitute a first step toward testing the cognitive validity of the directing rules underlying virtual camera control.

### 1.2 Mental Representation Richness

It is noteworthy that the literature on spatial knowledge makes extensive use of natural or concrete objects and environments (Golledge, 1987). However, one may wonder if these findings generalize to VEs containing artificial objects, for which one has no knowledge structures or sets of expectations based on past experience. Therefore, in the present study, subjects acquired spatial knowledge of an unfamiliar environment containing unfamiliar objects. Is processing of visual features enough to remember the location of virtual objects in an unfamiliar room? Would giving them a name (phonemic encoding) or knowing their meaning (semantic encoding) and organization inside the room improve access to the virtual spatial knowledge? This paper addresses this effect of representation richness—or depth of processing (Lockhart & Craik, 1990)—on access to spatial knowledge for unfamiliar items.

### 1.3 Where Is What?

Close to the issue of imagery perspective is that of the information being accessed: objects versus locations. Determining how information on objects and locations is processed in order to produce a spatial representation became a main research topic both in the fields of spatial language and spatial cognition (Landau & Jackendoff,



**Figure 1.** Map view of the simplified version of the Yamantaka mandala palace used in the present study. Left: location of each item listed in Table 1. Right: direction of the different views generated from the 3-D model of the room.

1993). There exists enough evidence in both animal (Mishkin, Ungerleider, & Macko, 1983) and human (Haxby et al., 1991; Farah et al., 1988) research showing that information on objects (“what” system) and their locations (“where” system) is processed by two distinct neurological and functional pathways, and that rehearsal processes for locational information start earlier than those for objects (Mecklinger & Pfeifer, 1996; Lea, 1975). However, the interplay of imagery perspective, access to object identity and location, and depth of processing remains an unexplored issue in the literature that our study is now addressing.

#### 1.4 Overview of the Study

In order to examine the effect of the two imagery perspectives on the access to spatial knowledge issued from a desktop VE, we used a modified version of the experimental paradigm devised by Lea (1975) to perform a chronometric analysis of the method of loci. The method of loci (Yates, 1966) is a mnemonic which involves forming an image of a familiar room or other space and imaging the to-be-remembered objects or

items each in a specific location. At the time of recall, the locations (loci) in the image are inspected, and the items incorporated in them are identified. Up to forty items may be perfectly retrieved after such a mental travel (Crovitz, 1971; Ross & Lawrence, 1968). In this study, *locus* refers to “vantage point” rather than to the identity of the locus as was the case in Lea’s study.

The environment that we used is a simplified version of the main room of a three-dimensional model of the mandala palace (Leidy & Thurman, 1997) of Yamantaka according to the Tibetan tradition (Dorje & Black, 1971), realized at the Program of Computer Graphics at Cornell University (<http://www.graphics.cornell.edu/~wbt/mandala/>). Unknown to our subjects, the objects contained in the room are 3-D representations of ancient Sanskrit syllables (Lantsa-Dewanagari script), each representing a deity concerned with either a specific state of mind or a natural element (such as anger and water) in accordance with the Buddhist tradition (Tucci, 1973; Bryant, 1992). The room is square with one gate in the middle of each wall (Figure 1). The items are organized spatially in a meaningful way, with those referring to natural elements located near the corners of the

**Table 1.** Structural, Phonemic, and Semantic Features of Each Item Contained in the VE Used in the Present Study as well as Its Location as Illustrated in the Figure 1 Map

Item									
Item location (cf. Figure 1)	1	2	3	4	5	6	7	8	9
Phonemic Feature	k.se	ja	ma	sa	me	do	da	ru	h-u.m
Semantic Feature	Mis-knowledge	earth	ego	water	attachment	fire	jealousy	air	anger

room, and those relevant to states of mind at the level of each entrance (Table 1). Apart from this particularity, one may note that the meaning of the items has no spatial relevance of the kind that would help one predict where to find them in the room (like a soapdish that would be expected near a sink). This rules out any bias that cognitive schemata (Brewer & Treyns, 1981) may introduce for our understanding of basic spatial memory processes. This study included three necessary phases.

1. In the first phase, depending on the group, subjects learned the items (objects) at a structural level (shape and color) only, or in combination with either their phonemic or semantic features. Instead of inventing new names or meanings for each item, we decided to give the subjects the phonemic or semantic features of each item in the room according to the Tibetan tradition (Dorje & Black, 1971). A recognition task was used to check that subjects could correctly distinguish the structural features of the items from pseudo-items.
2. In the second phase, the subjects learned the locus-item associations, that is, where each item was located inside the room. A cued-recall task was used to test the locus-item learning.
3. In the third phase, from a starting locus (vantage point) displayed on the computer screen, subjects were asked to scan mentally the emptied room in order to retrieve either the  $n_{th}$  locus or the  $n_{th}$  item. The task was performed either from the center of

the room using first-person imagery, or from its periphery with the help of third-person imagery.

The hypotheses of this study were as follows.

1. The recall of the location of each item in the room should improve with depth of processing of the locus-item association.
2. Subjects would access the locus information (“where”) more rapidly than its associated item (“what”).
3. Adopting a first-person imagery would lead to greater scanning times than with a third-person imagery, because the missing visual scene surrounding the subjects would have to be mentally reconstructed.

## 2 Phase 1: Learning Unfamiliar Virtual Objects

### 2.1 Purpose

As a prerequisite for the VE memorization, participants learned each item (object) that the room contained independently from the virtual room itself. This phase included two parts: a *learning stage* (during which subjects memorized each item (object) displayed in a prototypical orientation and then visualized it mentally), and a *test stage* (in which subjects decided if the item shown to them was one they had learned or a distractor

(pseudo-item)). Depending on the group, subjects also learned the name or meaning associated with each item. Close examination of the mental rehearsal duration after each item presentation in the learning stage, as well as recognition latencies in the subsequent recognition task, allowed us to evaluate the difficulty associated with the encoding of unfamiliar items.

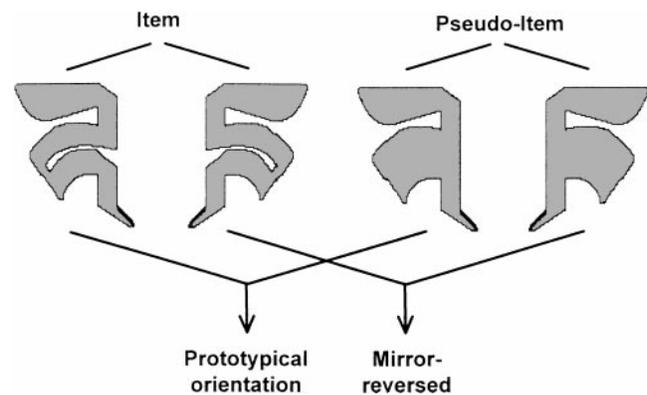
## 2.2 Method

**2.2.1 Participants.** Forty-eight subjects volunteered to participate in the study (24 men and 24 women) between 18 and 45 years of age ( $M = 29.7$ ;  $SD = 6.6$ ). They were either colleagues, lab staff, or visitors. All subjects had normal or corrected-to-normal vision and were naive to the purpose of the experiment and unfamiliar with Tibetan symbology. The subjects were distributed equally in three groups (three levels of encoding) with an equal number of males and females in each group.

**2.2.2 Experimental Setting.** The experiment took place in a quiet room under very dim lighting conditions. A personal computer with a NEC 5FG 17 in. graphic color monitor (VGA interface with a resolution of  $640 \times 480$  pixels) was used to present the stimuli and to record the responses. This phase as well as phase 2 and 3 were generated and monitored using ERTS-VIPL (BeriSoft Coop.), a PC-compatible software package that allows development and performance of psychological experiments (Beringer, 1994a, 1994b). The subjects sat comfortably in front of the display. The monitor was placed 80 cm away from the subject at eye level. Responses were recorded via the computer mouse. Text instructions (written in white characters) as well as colored stimuli (see Appendix A for their RGB values) were displayed against a dark background.

**2.2.3 Stimuli.** The items displayed to the subjects are presented in Table 1 in their prototypical orientation, together with their name and meaning. Figure 2 shows an example (item 7 of Table 1) of how distractors were obtained from the nine items, by filling an empty surface.

In the learning stage, items were displayed in their



**Figure 2.** Phase 1: the four versions of an item (here, item 7 in Figure 1). Its corresponding pseudo-item was obtained by filling the central part of the item.

prototypical orientation (unknown to our subjects, as readable Sanskrit characters), whereas in the test stage, both items and distractors were displayed in either the prototypical orientation or mirror-reversed.

## 2.2.4 Procedure.

**2.2.4.1 Learning Stage.** After a fixation point and warning signal were presented for one second, an item was displayed for thirty seconds in the center of the screen. Depending on the subject group, the item appeared alone, or was accompanied below and between parentheses (see Table 1) by either its name (phonemic feature, transliterated from Sanskrit) or its meaning (semantic feature). The subject was instructed to memorize the structural features of the item (its shape and color), as well as its name or its meaning, if present. The subjects who also encoded either the name or the meaning of the item were told that this supplementary information might be very important in subsequent phases of the study (phase 2 and 3). The subject was further invited, during each item presentation, to imagine what the item would look like if seen from the other side of the computer screen (that is, mirror-reversed), as well as to consider it as a three-dimensional object. After the item disappeared from the screen, text was displayed instructing the subject to close his/her eyes and to visualize the item both in its prototypical and mirror orientation, and also its accompanying name or meaning, if pre-

sented. Once the subject had mentally rehearsed the item long enough to feel confident in knowing it perfectly, he/she pressed a mouse button in order to trigger the display of the next item. If it was not the first item, the subject was also required to mentally visualize each of the previous items before proceeding to the next one. The items appeared in a pseudo-random order that was different for each subject.

**2.2.4.2 Test Stage.** Once the nine items had been sequentially displayed and rehearsed mentally, a recognition task verified that the items had been learned correctly. After presentation of a fixation point accompanied with a warning beep for one second, either an item or a distractor (pseudo-item) was displayed for three seconds in either its prototypical orientation or mirror-reversed (Figure 2). The screen was then blackened for seven seconds. The subject indicated with the mouse if a learned item had been displayed (left button) or not (right button). The response could be given during either the display of the test item or the subsequent black-screen display. The subject was informed if the allowed response time was exceeded or if the response was inaccurate. The next trial started automatically after the subject's response or the warning message. Therefore, the subject was forced to value speed (although instructed to be accurate). The test stage was preceded by three warm-up trials, which were not included in the data analysis.

Depending on the learning-stage group, if the item had been learned alone, both the items and pseudo-items were also displayed alone. In the other cases, if the items were learned either with a name or a meaning, this information also appeared below the test stimulus in the recognition task, whatever its orientation. If a pseudo-item was displayed, the name or the meaning of the item it originated from was displayed, too. The instructions insisted that the recognition judgment concerned the shape of the item whatever its orientation.

The subjects made a decision for each of the nine items as well as their corresponding pseudo-items, displayed in two possible orientations (prototypical or mirror-reversed), that is, 36 trials presented in a pseudo-random order. Each wrong or unanswered trial was repeated until the subject gave a correct answer to a

maximum of 72 trials in total, in order to avoid over-learning.

## 2.2.5 Experimental Design.

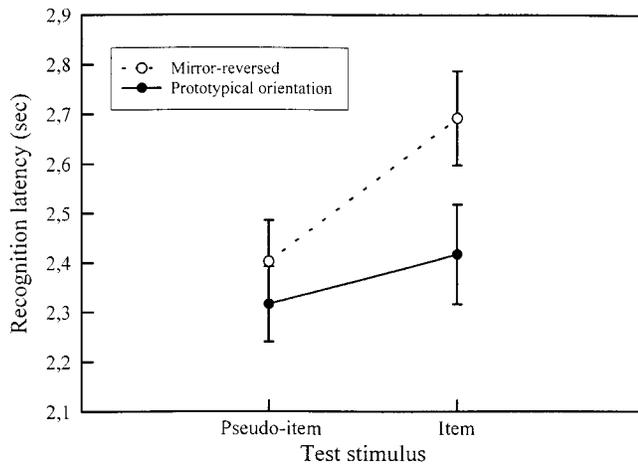
**2.2.5.1 Learning Stage.** We used a  $9 \times 3$  design in which item (either item type or trial number) was manipulated within-subject, and breadth of encoding (structural only, structural and phonemic, structural and semantic) was considered as a between-subjects factor. The dependent variable was the mental rehearsal duration which started after the disappearance of the item and ended with the push of a mouse button.

**2.2.5.2 Test Stage.** We used a  $2 \times 2 \times 3$  design in which item status (item versus pseudo-item) and item orientation (learned orientation versus mirror-reversed) were manipulated within-subject, whereas initial level of encoding at the learning stage (structural only versus structural and phonemic versus structural and semantic) was considered as a between-subjects factor. The dependent variables were recognition latency (time between the start of item display and the answer) and score.

## 2.3 Results

**2.3.1 Learning Stage.** The analysis of variance (ANOVA) of the mean duration of mental rehearsal subsequent to the item display showed neither an effect of the initial level of encoding [ $F(2, 45) = 1.34, p > 0.10$ , N.S.] nor of item type [ $F(8, 360) = 1.43, p > 0.10$ , N.S.]. However, we found a significant [ $F(8, 360) = 8.94, p < 0.0001$ ] increase of mental rehearsal duration with the trial number, reaching a plateau after the fifth item displayed regardless of its type.

**2.3.2. Test Stage.** In addition to measuring recognition latencies, we measured recognition score once corrected for guessing by subtracting the proportion of false alarms from that of hits (Snodgrass & Corwin, 1988). A corrected recognition score of 0% indicates a chance level. (Hereafter, for the sake of accuracy, when speaking of recognition performance or recognition score, we refer to the corrected recognition score). Nei-



**Figure 3.** Phase 1: interaction effects ( $\pm$ SEM) of item orientation  $\times$  item version on recognition latency.

ther recognition latency nor score was affected by initial encoding level [ $F(2, 45) < 1$ , N.S.]. However, significantly [ $F(1, 45) = 17.53$ ,  $p = 0.0001$ ] longer recognition latencies were found for the items ( $M = 2.57$  sec; S.D. = 1.08) than for the pseudo-items ( $M = 2.36$ ; S.D. = 1.00). This result suggests that subjects verified first if a distractor was displayed, and then processed the structural information further in order to double-check that it was an item rather than a distractor. Moreover, the results show that recognition latencies are significantly [ $F(1, 45) = 16.31$ ,  $p = 0.0002$ ] higher for the mirror-reversed stimuli ( $M = 2.53$ ; S.D. = 1.05) than those presented in a prototypical orientation ( $M = 2.38$ ; S.D. = 1.03). Along the same lines, items [ $F(1, 45) = 20.5$ ,  $p < 0.0001$ ] were significantly better recognized in their prototypical orientation ( $M = 0.88$ ; S.D. = 0.09) than when mirror-reversed ( $M = 0.79$ ; S.D. = 0.13).

Finally, close examination of Figure 3 reveals a significant orientation  $\times$  test stimulus interaction on recognition latency [ $F(1, 45) = 9.76$ ,  $p = 0.003$ ], indicating that the stimulus orientation affected response latencies to items rather than pseudo-items.

## 2.4 Discussion

Results of phase 1 indicate that learning the unfamiliar virtual objects was globally difficult. The similar

rehearsal durations for each item type points to an equal difficulty in creating a distinctive or unique trace in memory (Moscovitch & Craik, 1976; Jacoby, 1974) because of their unusual shapes. This process was certainly cognitively very demanding. The plateau in mental rehearsal duration reached after the fifth item seems to indicate that a limit in working memory capacity has been reached early, given Miller's (1956) finding that short-term memory span is limited to seven (plus-or-minus two) chunks.

Although during the learning stage subjects were instructed to consider the items as three-dimensional objects and imagine them mirror-reversed, they acted during the test stage as if they had built a viewpoint-dependent representation of each unfamiliar object. Hence, an additional process of mental rotation (Shepard & Cooper, 1982) was required to recognize mirror-reversed items among distractors.

The fact that the recognition of the items among distractors during the test stage was not affected by initial encoding level is consistent with the transfer-appropriate processing view (Morris, Bransford, & Franks, 1977), according to which the structural level of encoding capitalized on by all the subjects during the learning stage was also the most relevant processing level for completing the test stage. However, we kept displaying the name and meaning of the item to the subjects who had encoded these characteristics in the learning stage to strengthen the memorization of these associated features.

Given the overall good recognition performance, at the end of this stage the items were supposed to be learned, at least at the structural level. Although the name or meaning of the items were not relevant to recognize them, this information became relevant when learning their location in the virtual room, as we will see in phase 2.

## 3 Phase 2: Learning the Virtual Room

### 3.1 Purpose

Here, we examine the effect of the level of cognitive elaboration that was used to encode the organiza-

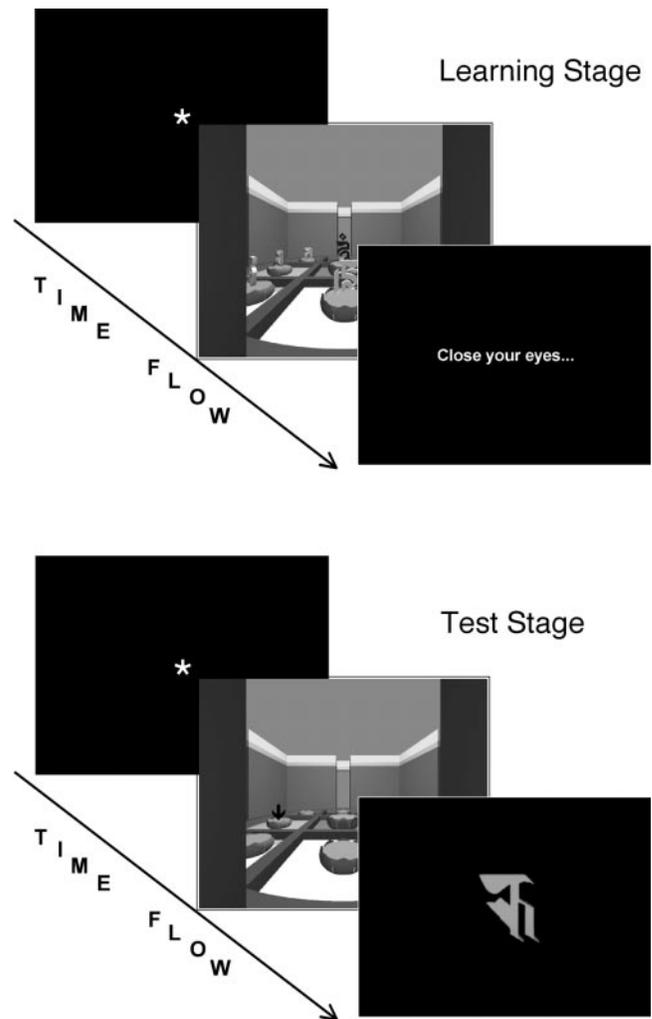
tion of the room on its memorization. On the one hand, we hypothesized that the recall performance of the objects' location would improve with the spread of encoding (from structural to semantic) of the spatial organization of the room. Learning the room on the basis of the functional organization of the items it contains would allow the chunking of spatial information and faster access to it, as compared to making use of the sequence of names of the items (phonemic rehearsal) contained in the room, or relying on their structural features (shapes and colors) only. In addition, we tested whether Roth and Kosslyn's (1988) finding that three-dimensional mental images are constructed in a near-to-far sequence applied to VE. Accordingly, recall of items located in the foreground of the visual scene would precede those located in the background.

### 3.2 Method

**3.2.1 Participants.** The same individuals that participated in phase 1 also participated in phase 2. The three groups were further split in half. Accordingly, subjects were distributed equally in six groups (three levels of encoding and two imagery perspectives) with an equal number of males and females in each group.

**3.2.2 Stimuli.** Different views were generated from a simplified version of the three-dimensional model of the Yamantaka palace room (Figure 1). The learning stimuli, consisting of one view per gate (so four gate views), were each generated with 100 degrees of viewing angle (Figure 4, top). In addition, for one group of subjects (first-person imagery), eight views were taken from the center of the room, whereas, for the other group (third-person imagery), eight views were taken from the periphery of the room (Figure 1). Each of these sixteen views had a viewing angle of sixty degrees.

For testing the room memorization, a view of the room was taken from the white gate, without any items. Nine exemplars of this view were generated, differing only by the presence of an arrow (cue) pointing down at an empty item location. Each white gate view contained an arrow that pointed to a different item location (Figure 4, bottom).



**Figure 4.** Phase 2: typical trial sequence for the virtual room (top)—learning phase (here, mental rehearsal after display of the white gate view), and (bottom) test phase (here, answer to the test item should be “No”).

**3.2.3 Procedure.** Once the characteristics of the items were memorized (phase 1), the learning of the location of the items in the room (locus-item association) started.

**3.2.3.1 Learning Stage.** After a fixation point and warning signal were presented for one second, the gate view taken from the white entrance was displayed for one minute. The subject was instructed to observe the room from this vantage point and to learn carefully how

the room is organized and where each item is located. Then, with eyes closed, the subject mentally visualized the room from the previously adopted vantage point and anticipated how the room would look from the next gate in the clockwise direction (that is, the yellow entrance). A typical trial sequence for one gate view is illustrated in Figure 4, top. Once the subject decided that he/she had formed a good representation of the room as it should appear from the next gate, he/she pushed a mouse button in order to actually observe the anticipated view of the room. The previous procedure was again applied for the current gate view (observation, mental rehearsal and anticipation, and mouse button push for the next gate view), as well as the next two gates. In other words, the subject made a clockwise tour of the room, jumping from one entrance to the other and anticipating the next vantage point each time, in order to elaborate a sufficiently complete representation of the room organization. Then, in order to improve the encoding of the room organization, depending on his/her group, the subject was shown eight successive views of the room (ten seconds per view) taken from either the center (first-person imagery group) or the periphery of the ring lying on the room floor (third-person imagery group). Each view was taken at the level of an item location (Figure 1, right). The sequential presentation of each view simulated a complete clockwise tour (view-point change) within the room.

Subjects were also encouraged to use all the features of the items for learning their location inside the room. Depending on the group, the room organization might be learned either on the basis of the structural features of the items (the color of the items is related to the color of the floor on which it is located), their phonemic characteristics (by articulatory rehearsal of the sequence of item names (Table 1)), or their meaning (natural elements are located in the corners of the room and states of mind are elsewhere).

**3.2.3.2 Test Stage.** A cued-recall task was initiated in order to test if subjects correctly learned the location of each item in the room. After a fixation point and warning signal were presented for one second, a view of the room taken from the white entrance was shown

without any items. However, the circular platforms that previously supported the items remained present and one had a downward pointing arrow above it. Each trial varied only by the item location pointed to by the arrow. The task was to recall the item whose location was cued by the arrow. The cued view was displayed for twenty seconds, then a test item was presented at the center of the screen for three seconds, followed by a black screen for seven seconds. A typical trial sequence is illustrated in Figure 4, bottom. The subject could interrupt the cued-view display at any time in order to see the test item by pushing a mouse button as soon as he/she recalled the item whose location was cued. The test item was one of the nine items learned by the subject and was displayed in the same way it had been learned (either alone or with its name or meaning depending on the group) in order for the subject to keep learning all the characteristics of the items. The subject indicated with the mouse if the test item belonged to the previously cued location (left button) or not (right button). The response could be given during either the display of the test item or the subsequent black-screen display. The subject was informed if the allowed response time was exceeded or if the response was inaccurate.

The subject gave his/her decision for the nine item locations, four times each (36 trials). For half of the trials, the test item was correct (answer should be “yes”), whereas, for the other half, an item not belonging to the cued location was randomly selected. The trials were presented in random order. Each wrong or unanswered trial was repeated until the subject gave a correct answer, up to a limit of 72 trials in total, in order to avoid over-learning. At the end of each trial, an instruction page invited the subject to push a mouse button when ready for the next trial. The test stage was preceded by three warm-up trials which were not included in the data analyses.

### 3.2.4 Experimental Design.

**3.2.4.1 Learning Stage.** We used a  $4 \times 3$  design in which the gate views and initial encoding level of items (structural only, structural and phonemic, structural and semantic) were considered as within-subject and be-

tween-subjects factors, respectively. The mental rehearsal duration for each gate view was measured and constituted the dependent variable.

**3.2.4.2 Test Stage.** We used a  $9 \times 3 \times 2$  design in which item location was manipulated within-subject, and initial encoding level of items (structural only versus structural and phonemic versus structural and semantic) as well as imagery group (first- versus third-person perspective) were considered as between-subjects factors. The dependent variables were recall latency, recognition time, and score.

### 3.3 Results

**3.3.1 Learning Stage.** The analysis of variance (ANOVA) showed a significant decrease of mean rehearsal duration ( $F(3,135) = 8.91, p < 0.0001$ ) after each gate view. This result suggests that the spatial knowledge of the room organization was more and more readily accessible from memory as learning evolved with the mental rehearsal process taking place after each gate view display. An alternative explanation would be that it reflects a shift in criterion (that is, subjects became satisfied with less, or more impatient, with additional trials), and not learning per se. There was no significant effect of the other factors on mental rehearsal duration.

**3.3.2 Test Stage.** In the cued-recall task, the ANOVA showed a significant effect of item location on both recall ( $F(8,336) = 19.68, p < 0.0001$ ) and recognition ( $F(8,336) = 7.23, p < 0.0001$ ) latencies. Table 2 shows the results for recognition score as a function of room organization (central item versus gate items versus corner items). The central item obtained the lowest latencies (recall:  $F(1,336) = 77.98, p < 0.0001$ ; recognition:  $F(1,336) = 40.38, p < 0.0001$ ) as well as the highest recognition score ( $F(1,336) = 10, p = 0.002$ ) as compared to all the other items taken together. The corner items showed greater latencies than the gate items for both recall ( $F(1,42) = 31.63, p < 0.0001$ ) and recognition ( $F(1,42) = 7.69, p = 0.008$ ) as well as a smaller recognition score ( $F(1,42) = 16.83,$

**Table 2.** Means and Standard Deviations for the Recall and Recognition Performance of Items as a Function of Room Topography in Phase 2

	Cued-recall latency (sec)		Recognition time (sec)		CR	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Central item	1.70	0.97	1.09	0.59	0.98	0.08
Gate items	2.55	1.75	1.41	0.87	0.92	0.11
Corner items	2.94	1.90	1.51	0.90	0.86	0.13

Note: CR, corrected recognition, difference between proportion of hits and false alarms. CR of 0% indicates chance performance.

$p = 0.0002$ ). These results suggest that the representation of the room built by the subjects is organized around the central and the gate items, which are de facto more readily accessible from memory. In order to rule out the confounding between semantic relatedness and room organization, we tested if any interaction existed between level of encoding and corner/gate effect. No such interaction was found on recall latency ( $F(2,42) = 2.38, p > 0.10, \text{N.S.}$ ) recognition time ( $F(2,42) < 1, \text{N.S.}$ ) and score ( $F(2,42) = 2.5, p > 0.09, \text{N.S.}$ ). In addition, the fact that this corner/gate effect was present when encoding the VE on the basis of its structural features (recall latency:  $F(1,14) = 24.18, p = 0.0002$ ; recognition score:  $F(1,14) = 6.44, p = 0.02$ ) definitely rules out any explanatory power of its confound with semantic relatedness.

In order to verify Roth and Kosslyn's (1988) finding that three-dimensional images are constructed in a near-to-far sequence, we considered the item locations as falling onto three successive depth planes, with respect to the vantage point of the white gate from which spatial judgments took place (Figure 4, bottom). Referring to Figure 1, items 1, 2, and 8 fall roughly on the foreground; items 4, 5, and 6 are in the background; and items 3 and 7 in the intermediate plane. Due to its peculiar status, we did not take into account the central item in this analysis. Table 3 shows the mean recall and recog-

**Table 3.** Means and Standard Deviations for the Recall and Recognition Performance of Items as a Function of Room Depth Planes from the White Gate in Phase 2

	Cued-recall		Recognition		CR	
	latency (sec)		time (sec)			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Foreground	2.53	1.70	1.44	0.88	0.88	0.14
Intermediate						
plane	2.77	1.90	1.43	0.90	0.92	0.15
Background	2.95	1.91	1.51	0.88	0.87	0.17

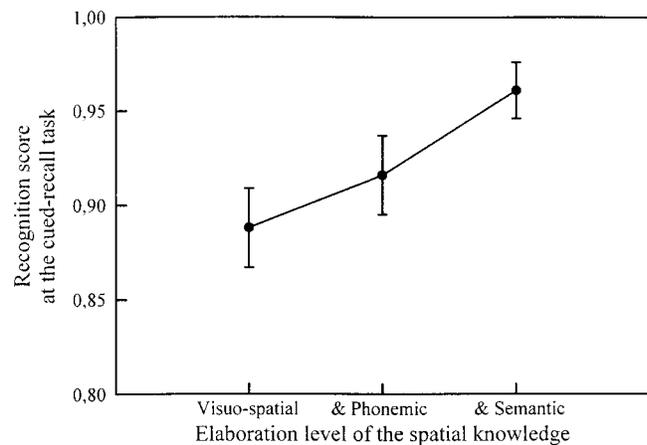
recognition performance for the items as a function of their locations in the depth planes.

As Roth and Kosslyn (1988) did, we found a significant increase of recall latencies with depth in the visual scene ( $F(2,84) = 10.38, p = 0.0001$ ) and a similar, although nonsignificant, tendency for recognition latencies ( $F(2,84) = 1.95, p > 0.10, N.S.$ ). Room depth planes had no effect on the recognition performance ( $F(2,84) = 1.88, p > 0.10, N.S.$ ).

With respect to the initial encoding level of the spatial knowledge, the ANOVA showed a significant improvement of recognition performance ( $F(2,42) = 4.2, p = 0.022$ ) as a function of level of encoding (Figure 5), but no effect on the time measurements. More precisely, the semantic level of cognitive elaboration of the room organization led to a significantly higher recognition score ( $F(1,42) = 7.4, p = 0.009$ ) with respect to the two other conditions. There was no effect of the other factors on latencies and recognition performance.

### 3.4 Discussion

Phase 2 has shown that spatial knowledge acquired from a VE, as for real environments, can be organized on the basis of information, such as distance and direction (Stevens & Coupe, 1978; McNamara, 1986). This was illustrated by the finding that the objects located near the entrances are accessed faster and more accurately than those lying near the corners of the room. In addition, we found that, from a given vantage point,



**Figure 5.** Phase 2: mean results ( $\pm$ SEM) for the recognition score (corrected for guessing) at the cued-recall of item location, as a function of the elaboration level of the spatial knowledge.

retrieval of memorized location of objects relies on processes similar to those involved in 3-D mental imagery generation. That is, information is accessed in a near-to-far sequence (Roth & Kosslyn, 1988).

Our finding of improved learning with level of encoding extends the hypothesis of functional grouping (Hirtle & Jonides, 1985) of spatial information, even to unfamiliar environments such as the VE that we used. Nonspatial information can also serve as an organizing principle. For example, people tend to group buildings by functions: that is, commercial buildings with other commercial buildings, and university buildings together themselves, despite the spatial scattering of the buildings (Hirtle & Jonides, 1985). Similarly, McNamara, Halpin, and Hardy (1992) demonstrated that nonspatial facts about objects are integrated in memory with spatial knowledge to facilitate location judgments.

In summary, VE information richness—may it be at a structural (gate/corner effect) or representational (level of encoding effect) level—improves the cognitive collage (Tversky, 1993) participating to the construction of a VE mental model. Also, if information about nonvisible objects from a given vantage point needs to be retrieved from the VE mental model, it will follow the rules of 3-D mental imagery (Roth & Kosslyn, 1988). The next phase investigated the cognitive processes involved in cognitive repositioning and access to object information

associated with the new vantage point, following first- and third-person imagery.

#### 4 Phase 3: Anticipating a New Vantage Point or Its Associated Object with Either a First- or a Third-Person Imagery

##### 4.1 Purpose

In his chronometric analysis of the method of loci, Lea (1975) showed that this mnemonic system involves two measurable mental processes: an iterative MOVE process that generates the successive loci, and a RETRIEVE process that acts at the end of the scanning process and retrieves the item associated with the locus. Therefore, access to the  $n_{th}$  locus precedes access to the  $n_{th}$  item. We wanted to test if this hypothesis applied to the context of VE with unfamiliar items and to examine if anticipation of the  $n_{th}$  vantage point would precede anticipation of its associated item. Indeed, in our study, *locus* refers to vantage point rather than to the identity of the locus, as was the case in Lea's study. Moreover, we compared the effect of two imagery perspectives on the cognitive exploration of a VE. Cognitive repositioning inside the virtual room was performed either with a first- or a third-person imagery.

##### 4.2 Method

**4.2.1 Participants.** The same individuals that participated in phase 1 and 2 also participated in the subsequent phase 3.

**4.2.2 Stimuli.** Different views of the room were again generated from its three-dimensional model (Figure 1). For the first-person and the third-person imagery groups, the same views as in phase 1 were taken from the center and the periphery of the room, respectively, but without the items in the room.

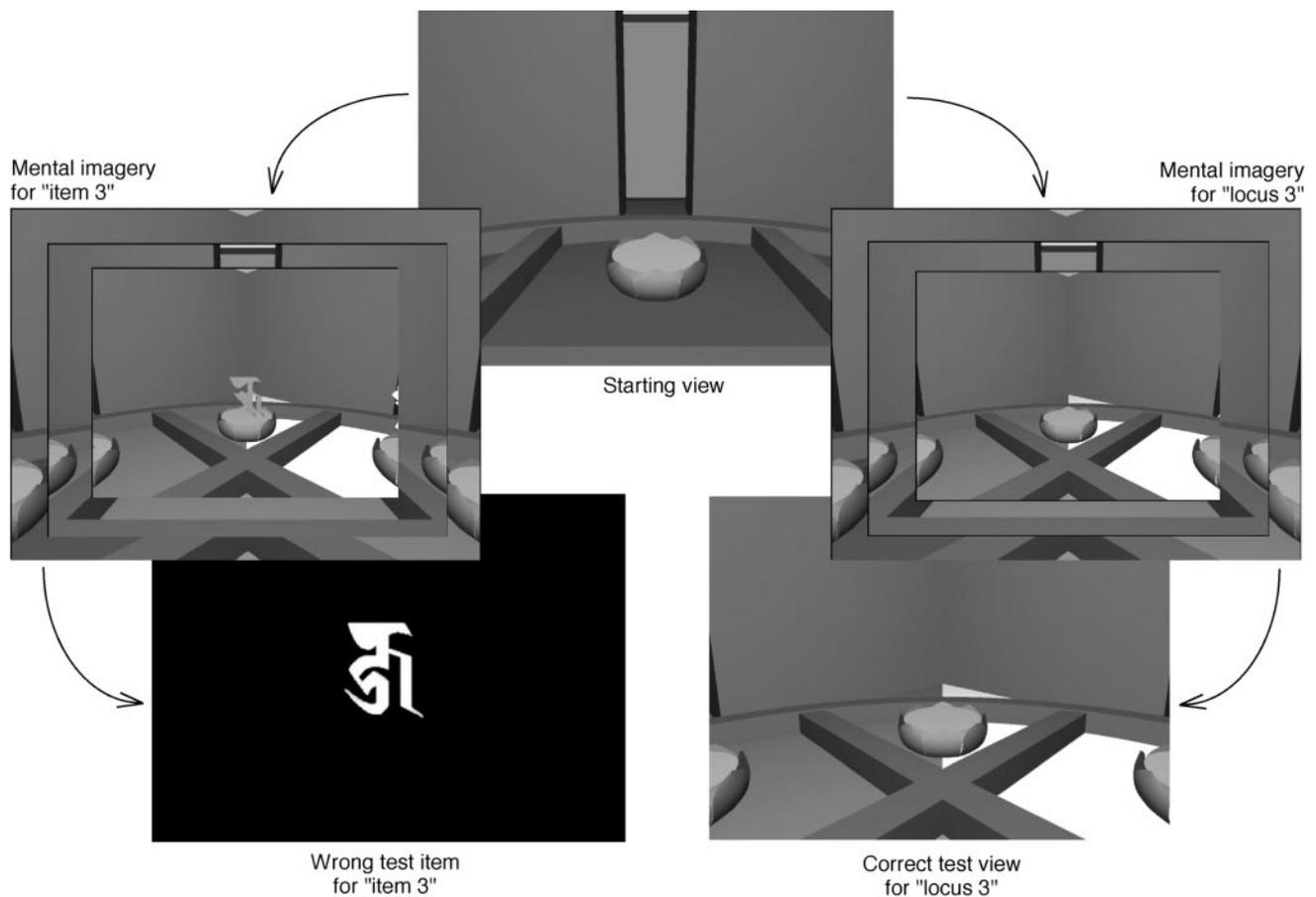
**4.2.3 Procedure.** After a fixation point and warning signal were presented for one second, a question was displayed on the monitor for three seconds.

The subject was asked to anticipate either the  $n_{th}$  vantage point (locus) or the  $n_{th}$  object (item). The value of  $n$  could vary from 0 to 7. The current vantage point was always 0 following an egocentric reference frame, and the to-be-imagined scanning direction in the room was always clockwise.

Figure 6 and 7 illustrate the stimuli and hypothesized imagery processes involved in the two imagery perspectives and searched information condition that will be discussed below. The subject was first informed what he/she was looking for (locus or item). Then, an initial vantage point (locus 0) at the level of an item location was displayed for an unlimited observation period. On the basis of the structural features (color and heading direction) of this starting view and the spatial knowledge acquired in phase 1, the subject mentally explored the empty environment clockwise until he/she reached the location corresponding to the question. The first-person imagery group viewed the room from its center, looking to the periphery, and imagined a self-rotation in place (Figure 6). The third-person imagery group viewed the room from its periphery, looking toward its center, and imagined both a translation and rotation of their view-point (Figure 7). Then, the subject pushed a mouse button in order to see either the test locus or the test item (scanning time). The subject indicated whether the test stimulus corresponded to the to-be-imagined vantage point or object (recognition time). At the end of each trial, an instruction page invited the subject to push any mouse button when ready for the next trial. This phase was preceded by six warm-up trials, which were not included in the data processing.

Phase 1 and 2 were learning phases in order to build a spatial knowledge of the unfamiliar VE. Results of phase 2 showed that the subjects who learned additional phonemic or semantic features included them in their spatial knowledge of the room organization. Therefore, in order to avoid overlearning and confusion with the scanning task, for the "item  $n$ " question, the test item appeared alone for all the subjects (that is, without any information regarding its name or meaning).

**4.2.4 Experimental Design.** All the subjects underwent 64 trials with a pause every sixteen trials.



**Figure 6.** Phase 3: example of stimuli and hypothesized mental imagery content in the first-person imagery condition, when looking for an object and anticipating a new vantage point. Locus 0 is the starting view.

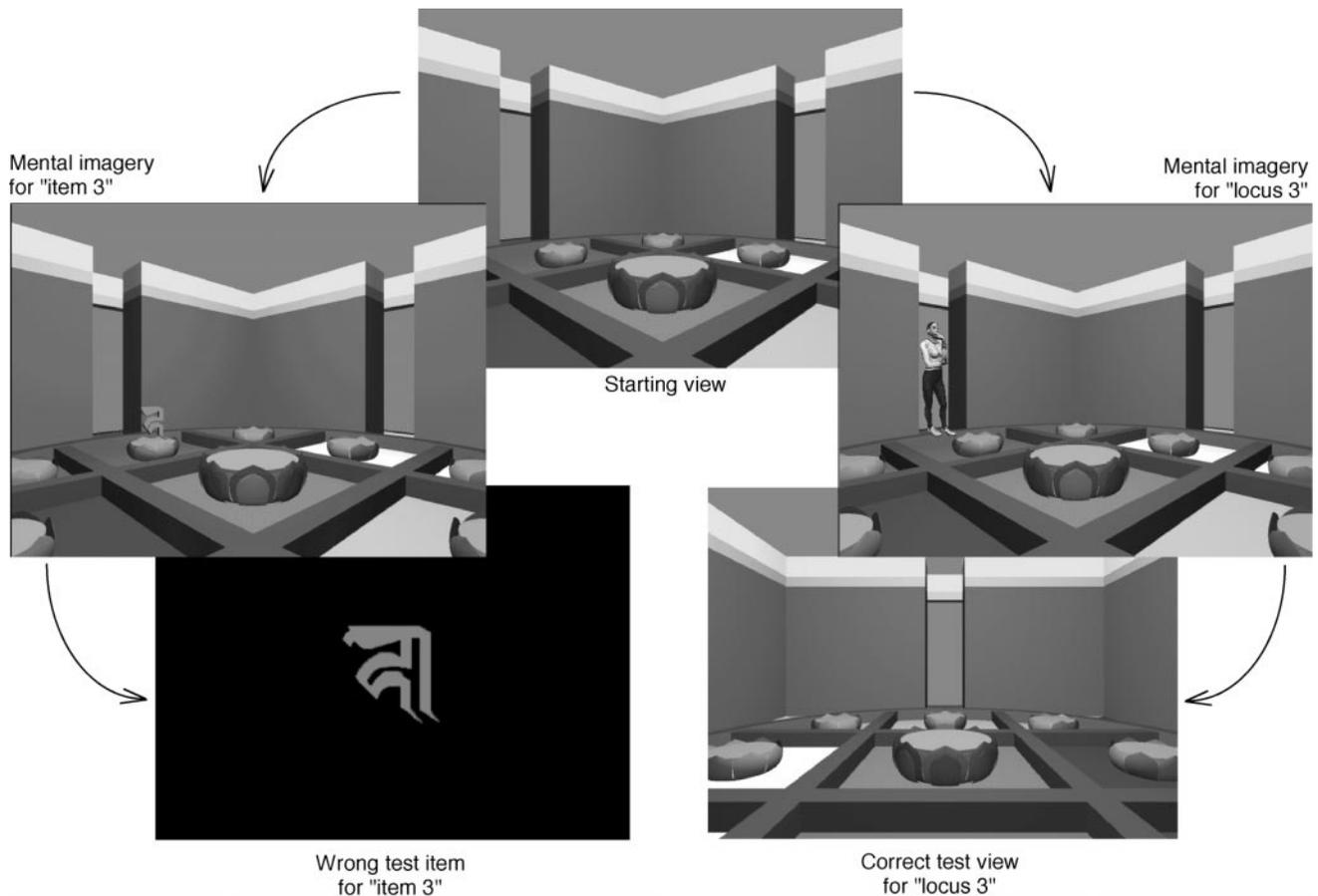
Each of the eight vantage points either in the center (first-person imagery group) or at the periphery (third-person imagery group) of the room were tested with each possible  $n_{th}$  locus scanned ( $n = 0$  to 7). A Latin square design was used to assign a  $n_{th}$ -locus question to half of the trials, and the  $n_{th}$ -item question to the other half, as well as to give an incorrect test stimulus for the recognition task in half of the trials, and a correct one for the other half.

In summary, the number of scanned loci and the searched information (locus versus item) were manipulated within-subject. The initial encoding level of items (structural only versus structural and phonemic versus structural and semantic) as well as imagery perspective (first- versus third-person) were considered as between-

subjects factors. The dependent variables were scanning time (subject-controlled SOA between initial locus and test stimulus), as well as recognition time and score for the test stimulus. In addition, in order to test the effect of the organization of the room on access to spatial knowledge, we treated the location of the test stimulus (close to a gate or a corner) as a supplementary experimental factor for the data analysis.

### 4.3 Results

**4.3.1 Scanning Process.** An ANOVA of scanning times showed significantly higher values ( $F(1,42) = 23.91, p < 0.0001$ ) for access to the  $n_{th}$  item ( $M = 7.81$ ; S.D. = 5.77 sec.) than for the  $n_{th}$  locus



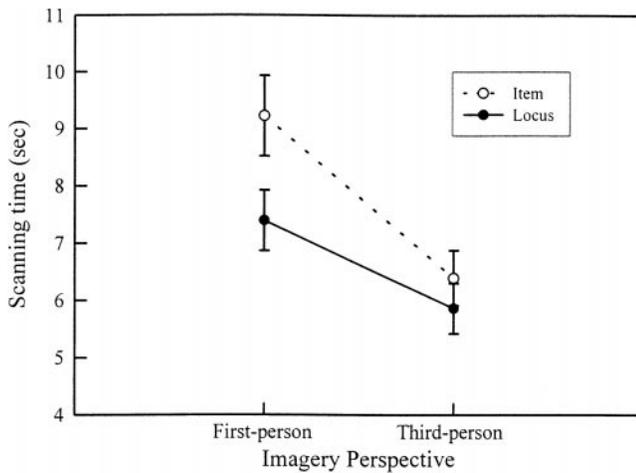
**Figure 7.** Phase 3: same as Figure 6 but for the third-person imagery condition (via an imaginary avatar when anticipating a new vantage point).

( $M = 6.63$ ;  $S.D. = 5.14$  sec.). The analysis of variance showed that adopting a first-person imagery ( $M = 8.31$ ;  $S.D. = 6.57$  sec.) was significantly more time consuming ( $F(1,42) = 8.27$ ,  $p = 0.006$ ) than with third-person imagery ( $M = 6.13$ ;  $S.D. = 3.86$  sec.).

Furthermore, there was a significant interaction ( $F(1,42) = 7.19$ ,  $p = 0.01$ ) between imagery perspective (first- versus third-person) and searched information (locus versus item) on scanning time (Figure 8). This reflects the finding that there were significantly higher latencies ( $F(1,21) = 27.22$ ,  $p < 0.0001$ ) for access to the  $n_{th}$  item ( $M = 9.22$ ;  $S.D. = 6.85$  sec) than to the  $n_{th}$  locus ( $M = 7.39$ ;  $S.D. = 6.14$  sec.) when adopting a first-person imagery, but no significant difference was found for the third-person imagery ( $F(1,21) = 2.58$ ,  $p > 0.10$ , N.S.). The ANOVA also showed a significant

increase of scanning time as a function of the  $n_{th}$  locus scanned ( $F(7,294) = 47.39$ ,  $p < 0.0001$ ). This effect varied significantly ( $F(7,294) = 17.53$ ,  $p < 0.0001$ ) as a function of the imagery perspective used to scan the VE mental model (Figure 9, top).

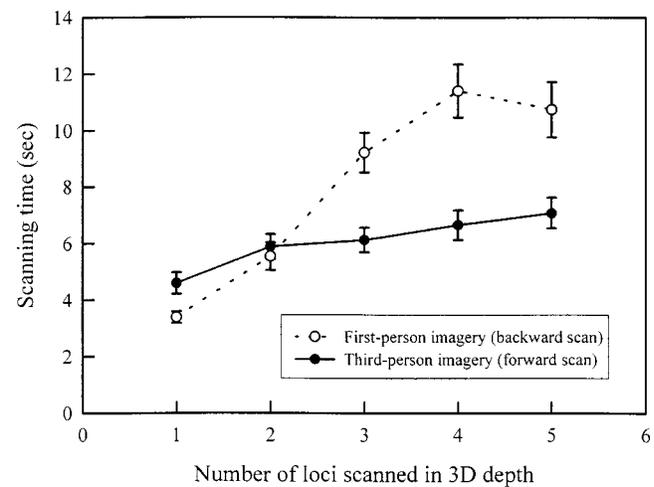
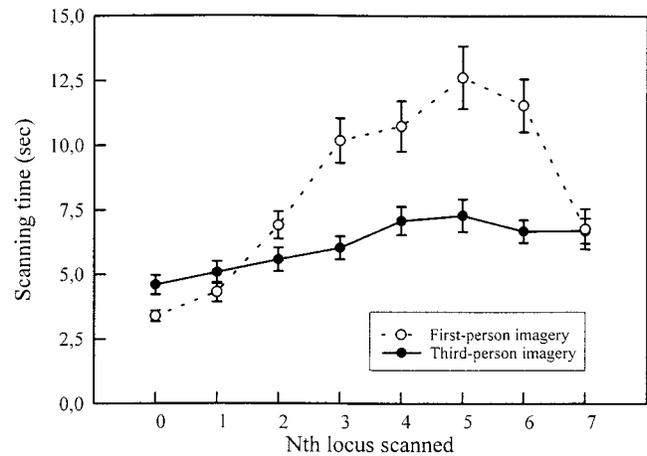
Figure 9 (top) shows, for the first-person imagery group, a decrease of scanning times after the fifth scanned locus. The  $n_{th}$  code (from 0 to 7) given to the subject corresponds to a clockwise tour in the room. However, the subjects did sometimes imagine a counter-clockwise movement instead: 0 (straightahead), 7, 6, 5, and so on, as suggested by post-experiment interviews and more objectively by the pattern of latencies illustrated in Figure 9 (top). Therefore, we decided to further analyze the scanning times as a function of the number of loci scanned relative to 3-D depth. In the



**Figure 8.** Phase 3: Interaction effects ( $\pm$ SEM) of searched information  $\times$  imagery perspective on scanning time.

first-person imagery condition, given that the array of objects is located around the observer, the scanning direction goes “backward” with respect to the observer’s viewing direction: starting straightahead toward either his left or his right side and then behind him. On the contrary, in the third-person imagery condition, the array of objects is located in front of the observer. Therefore, the scanning direction goes “forward” with respect to the observer’s viewing direction. Figure 9 (bottom) illustrates the scanning times obtained for both imagery perspectives (scanning directions) as a function of the number of loci scanned in 3-D depth. Accordingly, the data for locus 0 in Figure 9 (top) corresponds to one locus scanned in Figure 9 (bottom). Then, at least two loci should be scanned when the question is  $n = 1$  or 7 (that is, the starting locus  $n = 0$  and the next leftward or rightward locus), three loci for  $n = 2$  or 6, four loci for  $n = 3$  and 5, and finally five loci to reach  $n = 4$ .

Results of the analysis of variance show significant effects of the number of loci scanned in 3-D depth on scanning time which varies as a function of the imagery perspective. Table 4 shows the amplitudes of these observed effects as well as their corresponding  $F$  tests, and the Bayesian estimates (Bernard, 1994; Rouanet & Lecoutre, 1983; Rouanet, 1996) of the magnitude of the linear and nonlinear components. Globally, the observed



**Figure 9.** Phase 3: interaction effects ( $\pm$ SEM) on scanning time of (top) imagery perspective  $\times$   $n^{\text{th}}$  locus scanned and (bottom) imagery perspective  $\times$  number of loci scanned with respect to 3-D depth.

additional mean processing time per locus scanned is 1.37 sec. However, a trend analysis shows that the increase of scanning time with the number of loci scanned for the first-person perspective change is nonlinear (faster access to the opposite vantage point) and steeper than the linear trend of third-person imagery.

An interesting exception is the initial vantage point (locus 0), for which scanning times are significantly higher ( $F(1,42) = 7.88, p = 0.008$ ) for the third-person ( $M = 4.59; S.D. = 3.66$ ) than for the first-person ( $M = 3.39; S.D. = 3.10$ ) perspective group that we will discuss below.

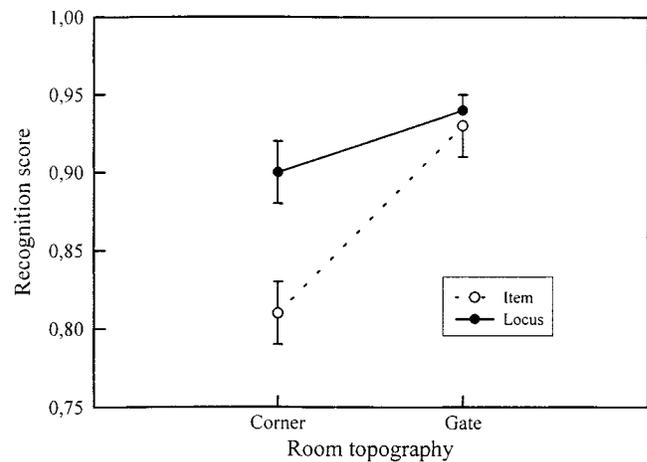
**Table 4.** Means ( $\pm$ SD) and Trend Analysis of the 3D Depth Effect on Scanning Time (sec) per Scanned Locus in Phase 3

	Observed effect	Bayesian estimate (at the 95% guarantee)	F Test
Globally			$F(4, 168) = 60.47^*$
linear component	1.37 ( $\pm 0.67$ )	1.10	$F(1, 42) = 82.44^*$
nonlinear component			$F(3, 126) = 16.03^*$
First-person imagery			$F(4, 84) = 47.58^*$
linear component	2.20 ( $\pm 0.95$ )	1.67	$F(1, 21) = 57.21^*$
nonlinear component			$F(3, 63) = 21.45^*$
Third-person imagery			$F(4, 84) = 15.91^*$
linear component	0.54 ( $\pm 0.11$ )	0.39	$F(1, 21) = 43.97^*$
nonlinear component			$F(3, 63) = 2.38 [p > 0.07, N.S.]$

Note— $F$  tests with  $*p < 0.0001$

**4.3.2 Recognition Process.** An ANOVA on recognition times showed significantly higher values ( $F(1,42) = 5.36, p = 0.026$ ) for the third-person ( $M = 1.78$ ;  $S.D. = 1.27$ ) than for the first-person ( $M = 1.40$ ;  $S.D. = 1.12$ ) perspective group. The ANOVA also showed significantly higher recognition scores ( $F(1,42) = 12.2, p = 0.001$ ) when looking for the  $n_{th}$  locus ( $M = 0.94$ ;  $S.D. = 0.07$ ) rather than the  $n_{th}$  item ( $M = 0.89$ ;  $S.D. = 0.10$ ), although the recognition times did not differ ( $F(1,42) < 1, N.S.$ ). Additionally, there was a significant interaction between searched information (locus versus item) and perspective group on recognition times ( $F(1,42) = 19.28, p = 0.0001$ ), whereby recognition times did not differ for items in both imagery perspectives ( $F(1,42) < 1, N.S.$ ), but took significantly longer times ( $F(1,42) = 18.05, p = 0.0001$ ) when checking the new vantage point (test view) in the third-person ( $M = 1.96$ ;  $S.D. = 1.32$ ) rather than first-person imagery ( $M = 1.29$ ;  $S.D. = 0.95$ ).

**4.3.3 Room Topography.** The reader may remember that, in phase 2, recall latencies for gate items were smaller than for those close to the corners of the room. Similarly, in phase 3, access to the  $n_{th}$  searched information close to the gates ( $M = 6.46$ ;  $S.D. = 5.28$ ) was significantly faster ( $F(1,42) = 88.88, p < 0.0001$ )



**Figure 10.** Phase 3: interaction effects ( $\pm$ SEM) of room topography  $\times$  searched information on recognition score.

than for those lying near the corners of the room ( $M = 7.97$ ;  $S.D. = 5.61$ ). The same significant difference was also found for the recognition time ( $F(1,42) = 41.71, p < 0.0001$ ) as well as a significantly higher recognition score ( $F(1,42) = 29.7, p < 0.0001$ ) for gate items ( $M = 0.95$ ;  $S.D. = 0.07$ ) rather than corner items ( $M = 0.87$ ;  $S.D. = 0.10$ ). In addition, we found a significant interaction ( $F(1,42) = 6.38, p = 0.015$ ) between searched information (locus versus

item) and room topography (corner versus gate) on recognition performance (Figure 10).

There was no significant effect of the other factors on times and recognition performance.

#### 4.4 Discussion

All the findings in phase 3 suggest that cognitive repositioning is governed by mental imagery processes at a structural level only. For instance, there was no effect of initial level of encoding on the anticipation of a new vantage point or its associated object. Instead, access to VE spatial knowledge was constrained by the room topography: The objects located near the entrances of the room were localized more rapidly and accurately than those near the corners. This result (Figure 10) suggests that subjects used a two-step rule in order to recall the item locations: the item facing a gate has the same color as the floor on which it is standing, and the next clockwise item (although standing between two floor colors) has the same color as its preceding item (counterclockwise direction). Accordingly, the entrances acted as retrieval cues for reconstructing the room environment.

Access to the new vantage point preceded access to its associated item. This was mainly true when adopting a first-person perspective for exploring the environment. This finding extends those of the chronometric study of the method of loci by Lea (1975) for two reasons. First, in this study *locus* refers to vantage point rather than to the identity of the locus. Second, access to the object was affected by the imagery perspective used for the cognitive repositioning.

Apart from counting the number of each location traversed to reach the new vantage point, cognitive repositioning involved different processes depending on the type of imagery perspective adopted. In the third-person perspective, the subject has a global view on the room which allows him to scan the available visual scene to extrapolate what would be the vantage point from the  $n_{th}$  locus and its associated item (Figure 7). On the other hand, in the first-person imagery, the subject has only a partial view on the room and must use an “image construction” strategy (Roth & Kosslyn, 1988) in order to retrieve the  $n_{th}$  locus or  $n_{th}$  item not visually available

(Figure 6). These different processes correspond to the “shift” versus “blink” transformations on mental image investigated by Kosslyn (1980, 1987), respectively for the third- and first-person type of cognitive repositioning. “Shift transformations involve altering an existing image, whereas blink transformations involve letting an initial image fade and then accessing stored information and generating a new image” (Kosslyn, 1987, p. 166). Our results are fully compatible with the higher processing load for blink transformation over shift transformation (Kosslyn, 1980; Kerr, 1993).

More evidence for the use of these distinct processes comes from close examination of Figure 9. For the initial vantage point (locus 0), scanning times are higher for the third-person than for the first-person perspective group. As illustrated in Figure 6, in the first-person imagery condition, the characteristics of the starting locus are clearly visible (that is, the color of the floor and whether it corresponds to a gate or a corner), whereas, in the third-person perspective, such characteristics are not visually available. This missing information has to be reconstructed mentally given what is available in the visual scene. On the contrary, for the next loci (from locus 2 to 7), third-person perspective subjects may explore the scene visually, whereas first-person perspective subjects have to reconstruct the missing visual scene (hence, the higher scanning times found in the later group for the next loci scanned).

We found a decrease of scanning time after four loci scanned (that is, when looking for either the locus or the item opposite to the starting vantage point) when using a first-person rather than a third-person imagery perspective. This effect is typical of perspective change studies (Hintzman, O’Dell, & Arndt, 1981; Rieser, 1989; Easton & Sholl, 1995) showing that targets either adjacent or opposite to the imagined orientation of the observer are more quickly accessed, given that the array of objects is located around the observer.

Finally, we found that imagery perspective affected the recognition process due to the richness of the visual scene at a structural level. The test view of the third-person perspective group, due to its global perspective on the room, contains more elements to check than the test

view of the first-person perspective group, which is a very partial view of the environment. Therefore it took more time to check the new vantage point (test view) in the third-person rather than first-person imagery, while the duration for checking the items was equivalent in both conditions.

The next section discusses the implications of the constraints introduced by first- versus third-person imagery and mental representation richness for cognitive repositioning inside a desktop VE.

## 5 General Discussion

### 5.1 Cognitive Elaboration of Spatial Knowledge

Our everyday spatial knowledge is inextricably elaborated through a conjunction of structural, phonemic, and semantic “spread” of encoding ( Craik & Tulving, 1975). VE environments are also defined both spatially and functionally (Colle & Reid, 1998). In this study, we tried to better evaluate the effect of representation richness by presenting the subjects with an unfamiliar environment for which encoding the different levels of processing could be manipulated. Our findings showed that, when available, subjects will use any opportunity to enrich their mental representation of a virtual space with additional information, either phonemic or semantic. Our subjects did use the semantic or functional information about the arrangement of objects to improve their cognitive elaboration of the location of the different items contained in an unfamiliar VE, even though this information was quite abstract (states of mind are in the corners and natural elements are near entrances). Therefore, it would be advisable to give virtual travelers the opportunity to structure the spatial information they encounter (such as the location of objects) even on abstract grounds. They would then chunk this information in the same way as chess players hold meaningful game configurations in memory (Ericsson & Smith, 1991) and have it more readily accessible. However, later on, it is mental

imagery processes that govern cognitive repositioning in VEs.

### 5.2 Searching for Items versus Locations

Our study is the first attempt to make a clear distinction between the mental anticipation of either the  $n_{th}$  point of view or its associated object. Indeed, one may question the likelihood of the locus-item association (*Locus* referring to a “place”) used by Lea (1975). In his Experiment 1, the loci are campus buildings and the associated items are either objects or animals, but—paradoxically—one of them is a mountain. More surprisingly, in his subsequent experiments, most of the loci are objects (Lea, 1975; Experiment 2 and 3).

Our results showed that there exists an interaction between imagery perspective and the searched information. (See response latencies in Figure 8.) Although Lea found higher response times when looking for the  $n_{th}$  item rather than the  $n_{th}$  locus, we found the same difference only when using a first-person perspective for exploring the environment. This additional processing time confirms that the locus must be reached before retrieving its associated item, and extends the finding of Lea by clarifying the constraints introduced by mental imagery on information retrieval.

### 5.3 First- versus Third-Person Imagery Perspectives

How an observer’s perspective on a virtual scene affects its representation is a major research topic among the cognitive issues in virtual reality (Wickens & Backer, 1995). Waller, Hunt, and Knapp (1998) argued that conscious effort and a widefield view (in terms of viewing angle) are necessary for acquiring a more accurate configurational knowledge of a VE. More than a question of viewing angle, our findings suggest that it is the amount of the virtual environment that is virtually viewable without moving the head which might affect the

construction and access to spatial knowledge. (By “virtually viewable,” we mean that there is more in a virtual scene than what is simply depicted.) Our contention is that it is the interplay of the content of the perceptual input and mental imagery-mediated memory processes that contributes to spatial updating and cognitive repositioning. Taking this view to the extreme, Thorndyke and Hayes-Roth (1982) suggest that, after extensive navigation experience, people can “look through” opaque obstacles in the environment to their destination without reference to the connecting route. They have built a survey knowledge from a perspective within, rather than above, the represented environment.

However, the drawback of relying on mental imagery is the increase of the cognitive workload of the observer, especially with first-person imagery. Therefore, giving the opportunity to switch between first- and third-person imagery might be of great benefit for the virtual traveler in order to anticipate new vantage points and appropriate action to take, as was also suggested by other authors in VE research (Bliss, Tidwell, & Guest, 1997; Ruddle, Payne, & Jones, 1997; Colle & Reid, 1998). Along these lines, Juurmaa and Lehtinen-Railo (1994) reported in their perspective-change study that subjects use this *doppelgänger* strategy to keep their own physical location in mind while they, in their imagination, shift to the new station. This strategy bypasses the cognitive cost of image-generation processes (Kosslyn, 1980, 1987; Roth & Kosslyn, 1988; Kerr, 1993) intrinsic to mental simulation of the visual effect of one’s navigation in a first-person mode of imagery.

The cognitive advantage of third-person over first-person imagery may be related to a misalignment effect between the physical viewing direction and the to-be-imagined orientation of the observer (Levine, Jankovic, & Palij, 1982; Presson & Montello, 1994) for the first-person rather than the third-person perspective conditions. Although instructed to keep facing the computer screen during the experimental trials, a few subjects in the first-person imagery group

found it necessary during the practice trials to close their eyes and move physically until facing the new vantage point in order to better visualize its content. This fits well with recent research in the field of human navigation showing that updating perspective change is facilitated by a physical movement of the subjects that should be consistent with the to-be-imagined perspective change, for both adults (Rieser, Guth, & Hill, 1986; Rieser, 1989; Young, 1989; Loomis et al., 1993; May & Wartenberg, 1995; May, 1996) and children (Huttenlocher & Presson, 1973; Rieser & Rider, 1991; Rieser, Garing, & Young, 1994). Similarly, Chance et al. (1998) showed that real motion improves spatial updating during first-person VE immersion. On the contrary, none of our third-person imagery participants felt the need to move physically in order to update the spatial surrounding.

Let us conclude with an anecdote from Tibetan tradition which tells us that updating perspective inside the “mind’s eye” mandala can be performed with the help of a real motion in the imaginary space, like in the ritual dance called ‘*cham*’: “the courtyard of the monastery becomes the mandala and all the dancers are transformed into the deities of that particular mandala” (Schrempf, 1994; p. 103). Hopefully, the secrets contained in this nice illustration of presence in a VE will lead to future understanding of the interplay between imagery perspective, movement, and representation richness.

## Acknowledgments

This study was supported by a grant from the Centre National d’Etudes Spatiales to M.-A. A. We would like to thank M. May, M. Wexler, and S. Wiener, as well as two anonymous reviewers for their comments on an earlier version of the paper.

Pema Losang Chogyen died of fibrosis in his chest on November 27, 1996, while a first draft of this article was being written. He was 39. Tibetan studies have lost one of their most valued contributors. We have lost a valued friend.

**Appendix A.** *The Following is a List of RGB Values that Describe the Colors Used in the Stimuli and Visual Scenes of This Study*

Stimuli	Colors		
	R	G	B
Item			
1, 2 (white gate)	255	255	255
3, 4 (orange gate)	241	157	27
5, 6 (red gate)	224	60	67
7, 8 (green gate)	21	201	81
9 (center)	79	0	183
Floor under			
item 1 (white)	255	255	255
3 (orange)	255	177	32
5 (red)	215	0	56
7 (green)	21	156	77
9 (blue-violet)	117	102	209
Lotus			
base (top)	255	132	89
base (side)	226	168	19
petals (rose)	223	95	128
(beige-brown)	182	127	67
(grey-blue)	90	100	152
Ring	99	52	214
Walls	99	52	214
(Medium band)	209	135	89
(Upper band)	255	220	22
Sky	127	127	204

## References

- Bares, W. H., Grégoire, J. P., & Lester, J. C. (1998). Realtime constraint-based cinematography for complex interactive 3D worlds. *Proceedings of the 10th National Conference on Innovative Applications of Artificial Intelligence*, 1101–1106.
- Bauman, S. (1996). *Third person 3D games feature: The question's not "Is your game 3D?" It's more like, "What type of 3D will it be?"*. [http://www.cdmag.com/action\_vault/3d\_games\_feature/page1.html]
- Beringer, J. (1994a). Software announcements—ERTS: A flexible software tool for developing and running psychological reaction time experiments on IBM PCs. *Behavior Research Methods, Instruments, & Computers*, 55 (4), 1.
- . (1994b). CiP94 Abstracts: ERTS-IPL: Tachistoscopic color image displays and accurate response registration on IBM PCs. *Psychology Software News*, 5, 37–38, CTI Centre for Psychology, University of York.
- Bernard, J.-M. (1994). Introduction to inductive analysis. *Mathématiques, Informatique, et Sciences Humaines*, 126, 71–80.
- Bliss, J. P., Tidwell, P. D., & Guest, M. A. (1997). The effectiveness of virtual reality for administering spatial navigation training for firefighters. *Presence: Teleoperators and Virtual Environments*, 6(1), 73–86.
- Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13, 207–230.
- Bryant, B. (1992). *The wheel of time sand mandala: Visual scripture of Tibetan Buddhism*. New York: HarperCollins Publishers.
- Chance, S. S., Gaunet, F., Beall, A. C., & Loomis, J. M. (1998). Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence: Teleoperators and Virtual Environments*, 7(2), 168–178.
- Colle, H. A., & Reid, G. B. (1998). The room effect: Metric spatial knowledge of local and separated regions. *Presence: Teleoperators and Virtual Environments*, 7(2), 116–128.
- Cox, M. V. (1977). Perspective ability: The relative difficulty of the other observer's viewpoint. *Journal of Experimental Child Psychology*, 24, 254–259.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268–294.
- Crovitz, H. F. (1971). The capacity of memory loci in artificial memory. *Psychonomic Science*, 24, 187–188.
- Dorje, G. L., & Black, S. M. (1971). A mandala of hJigs-bYed. *Oriental Art*, 17, 217–231.
- Drucker, S. M., & Zeltzer, D. (1995). CamDroid: A system for implementing intelligent camera control. *Proceedings of the SIGGRAPH '95 Symposium on Interactive 3D Graphics*, 139–144.
- Easton, R. D., & Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 483–500.
- Ehrich, V., & Koster, C. (1983). Discourse organization and sentence form: The structure of room descriptions in Dutch. *Discourse Processes*, 6, 169–195.

- Ericsson, K. A., & Smith, J. (Eds.) (1991). *Studies of Expertise: Prospects and Limits*. Cambridge: Cambridge University Press.
- Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, *20*, 439–462.
- Fishbein, H., Lewis, S., & Keiffer, K. (1972). Children's understanding of spatial relations: Coordination of perspectives. *Developmental Psychology*, *7*, 21–33.
- Golledge, R. G. (1987). Environmental cognition. In D. Stokols and I. Altman (Eds.), *Handbook of Environmental Psychology* (pp. 131–174). New York: John Wiley and Sons.
- Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., Herscovitch, P., Shapiro, M. B., & Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, *88*, 1621–1625.
- He, L., Cohen, M. F., & Salesin, D. H. (1996). The virtual cinematographer: A paradigm for automatic real-time camera control and directing. *Proceedings of ACM SIGGRAPH '96*, 217–224.
- Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology*, *13*, 149–206.
- Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory and Cognition*, *13*, 208–217.
- Huttenlocher, J., & Presson, C. (1973). Mental rotation and the perspective change problem. *Cognitive Psychology*, *4*, 277–299.
- Jacoby, L. L. (1974). The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning and Verbal Behavior*, *13*, 483–496.
- Juurmaa, J., & Lehtinen-Railo, S. (1994). Visual experience and access to spatial knowledge. *Journal of Visual Impairment & Blindness*, *88*, 157–170.
- Kerr, N. H. (1993). Rate of imagery processing in two versus three dimensions. *Memory & Cognition*, *21*, 467–476.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- . (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, *94*, 148–175.
- Landau, B., & Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*, 217–265.
- Lea, G. (1975). Chronometric analysis of the method of loci. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 95–104.
- Leidy, D. P., & Thurman, R. A. F. (1997). *Mandala: The architecture of enlightenment*. New York: Asia Society.
- Levine, M., Jankovic, I. N., & Palij, M. (1982). Principles of spatial problem solving. *Journal of Experimental Psychology: General*, *111*, 157–175.
- Light, P., & Nix, C. (1983). “Own view” versus “good view” in a perspective-taking task. *Child Development*, *54*, 480–483.
- Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology*, *44*, 87–112.
- Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. A. (1993). Nonvisual navigation by blind and sighted: Assessment of path integration ability. *Journal of Experimental Psychology: General*, *122*, 73–91.
- Mahoney, M. J., & Avenier, M. (1977). Psychology of the elite athlete: An exploratory study. *Cognitive Therapy and Research*, *1*, 135–141.
- May, M. (1996). Cognitive and embodied modes of spatial imagery. *Psychologische Beiträge*, *38*, 418–434.
- May, M., & Wartenberg, F. (1995). Rotationen und Translationen in Umräumen: Modelle und Experimente. *Kognitionswissenschaft*, *4*, 142–153.
- McNamara, T. P. (1986). Mental representations of spatial relations. *The Quarterly Journal of Experimental Psychology*, *15*, 211–227.
- McNamara, T. P., Halpin, J. A., & Hardy, J. K. (1992). The representation and integration in memory of spatial and nonspatial information. *Memory & Cognition*, *20*, 519–532.
- Mecklinger, A., & Pfeifer, E. (1996). Event-related potentials reveal topographical and temporal distinct neuronal activation patterns for spatial and object working memory. *Cognitive Brain Research*, *4*, 211–224.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review*, *63*, 81–87.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends In NeuroSciences*, *6*, 414–417.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Level of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.
- Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in

- recall. *Journal of Verbal Learning and Verbal Behavior*, 15, 447–458.
- Paniaras, I. (1997). *Design of virtual identities and behavior in cyber communities*. Paper presented at the 85th CAA Conference: Crossing the boundaries—Electronic Art Within and Without, New York. (<http://www.uiah.fi/~paniaras/caa2.htm>).
- Piaget, J., & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Kegan Paul.
- Presson, C. C., & Montello, D. R. (1994). Updating after rotational and translational body movements: Coordinate structure of perspective space. *Perception*, 23, 1447–1455.
- Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1157–1165.
- Rieser, J. J., Garing, A. E., & Young, M. F. (1994). Imagery, action and young children's spatial orientation: It's not being there that counts, it's what one has in mind. *Child Development*, 65, 1254–1270.
- Rieser, J. J., Guth, D. A., & Hill, E. W. (1986). Sensitivity to perspective structure while walking without vision. *Perception*, 15, 173–188.
- Rieser, J. J., & Rider, E. A. (1991). Young children's spatial orientation with respect to multiple targets when walking without vision. *Developmental Psychology*, 27, 97–107.
- Roehl, B. (1998). Internet multi-user worlds. *VR News*, 7(9), 10–15.
- Ross, J., & Lawrence, K. A. (1968). Some observations on memory artifice. *Psychonomic Science*, 13, 107–108.
- Roth, J. D., & Kosslyn, S. M. (1988). Construction of the third dimension in mental imagery. *Cognitive Psychology*, 20, 344–361.
- Rouanet, H. (1996). Bayesian methods for assessing importance of effects. *Psychological Bulletin*, 119, 149–158.
- Rouanet, H., & Lecoutre, B. (1983). Specific interference in ANOVA: From significance tests to Bayesian procedures. *British Journal of Mathematical and Statistical Psychology*, 36, 252–268.
- Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). Navigating buildings in “desk-top” virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied*, 3, 143–159.
- Schrempf, M. (1994). Tibetan ritual dances and the transformation of space. *The Tibet Journal*, 19, 95–120.
- Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press, Bradford Books.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34–50.
- Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, 10, 422–437.
- Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14, 560–589.
- Tucci, G. (1973). *The theory and practice of the mandala*. New York: Samuel Weisner.
- Tversky, B. (1993). Cognitive Maps, Cognitive Collages, and Spatial Mental Models. In A. U. Frank & I. Campari (Eds.), *Spatial Information Theory: A Theoretical Basis for GIS, International Conference COSIT '93*. Proceedings: Lecture Notes in Computer Science, 716, 14–24, Springer.
- Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperators and Virtual Environments*, 7(2), 129–143.
- Wickens, C. D., & Backer, P. (1995). Cognitive issues in virtual reality. In W. Barfield and T. A. Furness III (Eds.), *Virtual environments and advanced interface design* (pp. 514–541). Oxford: Oxford University Press.
- Yates, F. (1966). *The art of memory*. London: Routledge & Kegan Paul.
- Young, M. F. (1989). *Cognitive repositioning: A constraint on access to spatial knowledge*. Unpublished doctoral dissertation: Vanderbilt University.