SEPARATE LONG-TERM VISUAL AND SPATIAL REPRESENTATIONS

By

Christine Marie Valiquette

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
In partial fulfillment of the requirements
For the degree of

DOCTOR OF PHILISOPHY
in

Psychology

August, 2008

Nashville, Tennessee

Approved:

Timothy P. McNamara

Adriane E. Seiffert

John J. Rieser

René Marois

To Aiice, who believed in me, even when I did not

ACKNOWLEDGEMENTS

This work would not have been possible without the financial support of the National Institute of Health. I am especially indebted to Tim McNamara, my advisor and chairman of my Dissertation Committee, who fostered my independence and creativity, but insisted that I develop the discipline to support my intuitions with evidence.

I am grateful to all of those with whom I have had the pleasure to work during this and other related projects. Each member of my Dissertation Committee helped to make this project a success, by asking critical questions and contributing unique ideas. This project would not have been nearly as enjoyable without the contributions of Sara Delheimer, who was tremendously helpful with the preparation and conduct of the experiment, and wonderfully funny. With her help, this project often felt more like play than work.

No one has been more important to me in the pursuit of this project than my family. I would like to thank my parents, Al and Alice, my aunt Joan, my brothers and sisters, and nieces and nephews, all of whom provided much needed support and encouragement throughout this process.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Memory-based human navigation is composed of at least two components: determining one's location and determining the relative location of an unseen goal. The ability to recognize configurations of objects or landmarks plays a large role in determining one's current location, which is essential for staying oriented in, or reorienting with respect to, the environment. Determining the location of an unseen goal in relation to one's present location is most helpful when one's view is obstructed (e.g., by buildings) and for navigation that requires finding a novel route to a familiar location. Route following can be construed as alternation between determining one's present location and determining the relative location of a seen or unseen goal (i.e., the next decision point on the route).

Prominent theories of spatial memory and navigation do not distinguish between establishing one's present position and locating an unseen goal relative to that position in their analyses of long-term spatial memory (Easton & Sholl, 1995; McNamara, 2003; Mou, McNamara, Valiquette, & Rump, 2004; Shelton & McNamara, 2001; Sholl & Nolin, 1997; Wang & Spelke, 2000, 2002). Wang and Spelke claim that long-term spatial memory is composed of a viewpoint-dependent representational system and a geometry-based reorientation system. Multiple "viewer-centered" (Wang & Spelke, 2002) landmarks and scenes are stored in the viewpoint-dependent system, while the shape of the enclosing environment is stored in a separate geometry-based system. Spatial relations among objects in the environment are not represented in long-term memory. The viewpoint-dependent system is accessed for place recognition, which facilitates determining one's position. The viewpoint-dependent system also appears to support goal localization, as it is the only system proposed by Wang and Spelke that includes long-term representations of objects and landmarks.

In McNamara, Mou and colleagues' theory (McNamara, 2003; Mou, McNamara et al., 2004; Shelton & McNamara, 2001), an environmental subsystem represents enduring features of the environment. Intrinsic reference systems are used within the environmental subsystem to represent object-to-object spatial relations in an orientation-dependent manner. The intrinsic reference system used within a given environment is determined by how the environment is experienced. Egocentric experience, salient external frames of reference, and instructions have been demonstrated to be cues to intrinsic organization (Mou & McNamara, 2002; Shelton & McNamara, 2001; Valiquette & McNamara, 2007). The long-term memory for an environment is represented with respect to the intrinsic frame of reference selected upon the initial encounter with an environment unless a more salient frame of reference is subsequently experienced. Both determination of one's present position and goal localization are assumed to depend on spatial information housed within the environmental subsystem, although the process by which it is accessed has not been determined.

Sholl's theory (Easton & Sholl, 1995; Sholl, 2001; Sholl & Nolin, 1997) stipulates that spatial relations among objects are represented in an orientation-independent object-to-object reference system. Although it is not addressed by the theory, it is reasonable to assume that the object-to-object system is accessed to determine one's present location and locate an unseen goal.

Finally, Byrne, Becker and Burgess (Byrne, Becker, & Burgess, 2007) have proposed a model of spatial memory that encompasses findings from both cognitive psychology and neuroscience. The long-term representation of objects' locations is allocentric, and formed in the hippocampus. It consists of associations among egocentric spatial representations that are housed, short-term, in parahippocampal areas of the brain. The term short-term is not defined in their model, but their use of the term implies that the egocentric representations are maintained for a period of time without attention, making them longer lived than the egocentric representations proposed by McNamara and colleagues (McNamara, 2003; Mou & McNamara, 2002; Mou, McNamara et al., 2004) and Sholl (Easton & Sholl, 1995; Sholl, 2001; Sholl & Nolin, 1997), but not long term like the egocentric representations proposed by Wang and Spelke (2000, 2002).

It is possible that multiple egocentric, snapshot-like representations (Wang & Spelke, 2000, 2002) and a single allocentric, orientation-dependent object-to-object representation (McNamara, 2003; Mou et al., 2004) are maintained in long-term memory. Through visual matching, snapshot-like representations could facilitate place recognition used to determine one's present position, whereas an amodal spatial, object-to-object representation could facilitate locating an unseen goal based on recalled object-to-object relations. This dual-type representational system was proposed by Valiquette and McNamara (2007).

Three recent experiments provide support for such a dual-type system (Shelton & McNamara, 2004a, 2004b; Valiquette & McNamara, 2007). In each of the experiments, participants learned the identities and locations of objects in multi-object layouts. They then completed a scene recognition task and judgments of relative location (JRD) in a remote location. Scene recognition showed benefits for all experienced views, providing support for multiple snapshot-like representations. Judgments of relative direction showed benefits for headings corresponding to the dominant direction or axes of the intrinsic frame of reference, even when the dominant direction was never visually experienced (Shelton & McNamara, 2004b), supporting the notion of a single, amodal, object-to-object representation.

For example, Valiquette and McNamara (2007) conducted an experiment in which two views of a layout of objects were experienced through the same sensory modality, vision, for the same amount of time, with the same amount of emphasis placed on each during learning. The study procedure was identical to that used by Shelton & McNamara (2001, Exp 3). Participants studied the locations of seven objects placed on a square mat within a rectangular room from two viewing positions (see Figure 1). One position was aligned with the walls of the room and the edges of the mat. The other position was misaligned with both the walls of the room and edges of the mat. Order of study was counterbalanced across participants. Participants studied from each position until they could point to and name all of the objects with their eyes closed, resulting in approximately equal exposure to each view of the layout. After learning the objects' locations to criterion, they performed a scene recognition task and a JRD task in a remote location. Each scene recognition test trial required participants to determine if a picture presented on a computer screen was of the configuration of objects that they had studied, or a picture of the same objects occupying different locations (the distractor layout). Judgment of relative direction test trials were composed of the names of three of the objects in the layout and required participants to point to an object as though standing at a particular location within the layout; for example, "Imagine you are standing at the *book* facing the *wood*, point to the *clock*." Trials were presented as text on a computer screen, and participants responded by pointing in the appropriate direction using a joystick.

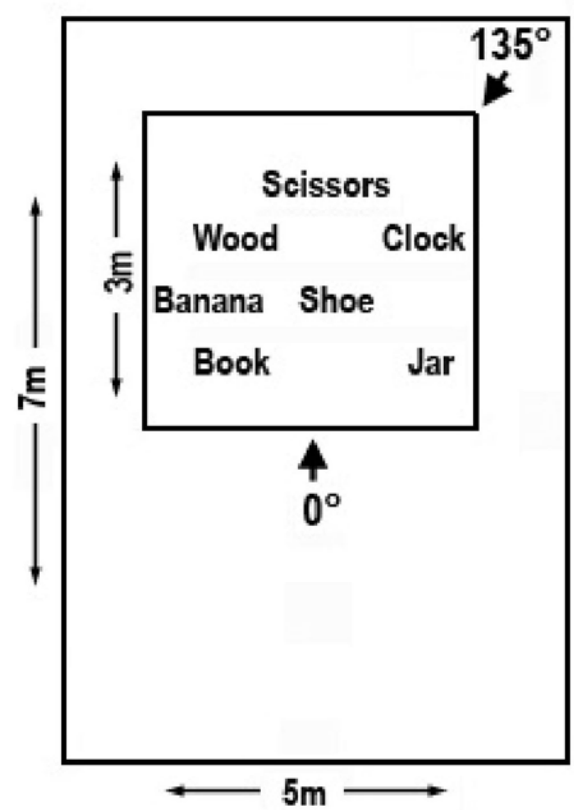


Figure 1. Diagram of the study layout used by Valiquette & McNamara, 2007.

The JRD results replicated Shelton and McNamara's (2001) results, with a benefit for the imagined heading congruent with the aligned study position and no benefit for the imagined heading congruent with the misaligned study position over novel imagined headings. In fact, performance was better at four novel headings (45°, 90°, 180°, and 270°, statistically significant for 90°) than at the imagined heading corresponding to the 135° study view. Performance was better on the non-zero-degree imagined headings aligned with the salient orthogonal axes of the room and mat (90°, 180°, 270°) than for misaligned imagined headings (45°, 135°, 225°, 315°), producing a saw-tooth pattern of results. The same saw-tooth pattern was found in the reaction time data.

The finding of no benefit for JRD trials in which the participant imagined facing a direction they faced while extensively studying the layout is highly counter-intuitive, but has been replicated numerous times (Mou & McNamara, 2002; Shelton & McNamara, 2001, 2004b; Valiquette & McNamara, 2007; Valiquette, McNamara, & Labrecque, 2007). These findings support McNamara and colleagues claim that long-term spatial memory consists of a single allocentric orientation-dependent representation, not necessarily based on direct perception or egocentric experience.

The saw-tooth pattern within the angular error results has been replicated numerous times as well (McNamara, 2003; McNamara, Rump, & Werner, 2003; Mou & McNamara, 2002; Valiquette et al., 2007; Valiquette, McNamara, & Smith, 2003). It indicates that participants retrieved inter-object spatial relations more easily when they imagined facing directions aligned with the walls of the study room. The saw-tooth pattern has been attributed to the organizational cues provided by orthogonal axes within the environment (walls of room, square mat), and to organizational cues provided by the intrinsic structure of the display itself (Mou & McNamara, 2002).

In contrast, the scene recognition results showed benefits for pictured headings corresponding to both study views. The saw-tooth pattern seen in the JRD results was not apparent in the scene recognition results, indicating that the representations accessed to perform the scene recognition task were not organized along axes congruent with the structure of the surrounding environment. Taken together, these two components of the results pattern support Wang and Spelke's (2000, 2002) idea of multiple egocentric snapshot-like representations.

The differing pattern of results from JRD and scene recognition led Valiquette and McNamara (2007) to conclude that both a single allocentric and multiple snapshot-like egocentric representations were housed in long-term memory. This can be illustrated by an example. A student who sits in the same seat every time he or she attends a class in a rectangular room in which the desks are arranged in rows and columns would be expected to form multiple representations of the space. In the single allocentric representation, the student's seat is represented with respect to the rows and columns of desks, and he or she is represented as an object within the arrangement of objects in the room. In the multiple snapshot-like representations, the relative locations of objects in the room depend upon the position from which the representation was encoded (i.e. from the door, from the podium, etc.). When the student is shown pictures of classrooms (i.e., scene recognition), he or she uses visual matching to the multiple snapshot-like representations to decide if the pictures presented are views of the classroom. When asked to imagine being in a specific position within the classroom and then point to an object in that room (i.e., JRD), the student accesses the allocentric representation to determine the object's location relative to the imagined position.

A recent article by Mou, Fan, McNamara, and Owen (2008) argues against the existence of long-term visual memories. Rather than separate visual and spatial representations, they propose two elements within a single amodal, spatial representation. According to their proposal, people represent the location of objects with respect to an orientation-dependent, intrinsic frame of reference and represent the direction from which they viewed the objects with respect to the same intrinsic frame of reference. This explanation allows for a single representation to support the viewpoint dependence found in scene recognition and orientation dependence (which may or may not correspond to viewpoint dependence) associated with JRD. In contrast to the single allocentric and multiple snapshot-like representations formed by the student in the above example, Mou et al.'s single representation contains the locations of objects in the room coded with respect to the axes highlighted by the rows and columns of desks. The location of the student's desk, the door of the room, and other positions from which the room was viewed are encoded with respect to the structure provided by the rows and columns of desks. Objects' locations are not represented with respect to the location from which they were viewed, rather the viewing location is represented with respect to axes congruent with the rows and columns of desks. The student accesses the same representation to determine if a picture is of the classroom (scene recognition) as to determine an object's location relative to an imagined position within the room (JRD). Snapshot-like representations of what the room looks like from the door, from the student's desk, etc., are not preserved in Mou et al.'s single representation. Therefore, a determination of whether or not a picture is of the classroom cannot be made using visual matching.

In Mou et al.'s (2008) experiments participants learned a layout of seven objects by viewing them from one location, while being instructed to learn the objects' locations in columns misaligned with the viewing location by 45° (see Figure 2a). Previous studies (Mou & McNamara, 2002) have shown that this method effectively leads participants to organize their mental representations of the layout along the intrinsic axes corresponding to the highlighted columns. Participants then performed a scene recognition task. In one experiment (Exp. 2) scene recognition was performed using test layouts consisting of six of the seven studied objects. As in previous experiments (Diwadkar & McNamara, 1997; Shelton & McNamara, 2004b; Valiquette & McNamara, 2007), performance was best for the heading from which the layout was viewed. Another experiment (Exp. 3) employed a scene recognition task in which triplets of objects were presented, rather than the standard presentation of entire layouts. There were two categories of trials: intrinsic, containing two objects in one of the columns that defined the intrinsic structure of the layout (indicated by rectangles in Figure 2a), and non-intrinsic, containing two objects in one column aligned with the viewing position (indicated by ellipses). Both trial types were presented at all headings (0° - 315°, in 45° increments). Intrinsic trials produced shorter reaction times than non-intrinsic trials (see Figure 2b). The results also revealed a separate benefit for the studied view, regardless of trial type.

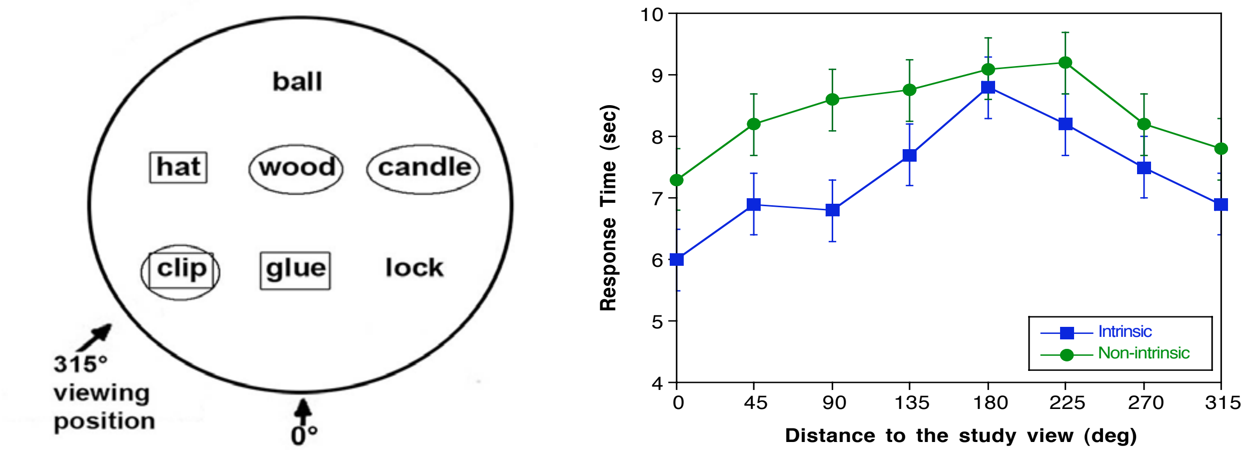


Figure 2. (a) Diagram of the layout used by Mou et al., 2008, rectangles indicate intrinsic triplets, ellipses indicate non-intrinsic triplets (b) response time on Mou et al.'s scene recognition target trials as a function of pictured heading and triplet type.

Mou et al. (2008) claim that the benefit for intrinsic trials over non-intrinsic trials demonstrates that snapshot-like, visual representations of the environment are not stored in long-term memory. Instead, they propose that the intrinsic structure of the environment is encoded with respect to the position from which it was experienced. They assert that, if snapshot-like representations of the layout were stored in memory, non-intrinsic triplets would be recognized equally as quickly as intrinsic triplets. According to Mou et al., the role of intrinsic structure in scene recognition has been overlooked previously because test trials have consisted of entire layouts of objects, in which the intrinsic structure is always available. Mou et al. did not address how the single representation that they propose would incorporate multiple views of a layout.

As discussed above, scene recognition has been proposed to access egocentric, snapshot-like, representations of the environment (Diwadkar & McNamara, 1997; Valiquette & McNamara, 2007; Wang & Spelke, 2000, 2002). It is unclear if egocentric means self-to-individual objects in a multi-vector manner (e.g., Sholl & Nolin, 1997), or self-to-array, which corresponds better intuitively with how snapshots are experienced, but has not been conceptualized in the spatial memory literature. The term egocentric will imply self-to-array relations for the purposes of the present discussion.

An allocentric representation of the environment appears to be accessed to localize an unseen goal (e.g., JRD; McNamara, 2003; Mou & McNamara, 2002; Valiquette & McNamara, 2007). Allocentric is also a somewhat ambiguous term. Wang and Spelke (2000, 2002) claim that an allocentric representation of the shape of the environment (but not inter-object relations) is stored in long-term memory, whereas others (Byrne et al., 2007; Mou, McNamara et al., 2004; Valiquette & McNamara, 2007; Waller & Hodgson, 2006) use the term to refer to object-to-object spatial relations.

The experiment presented here is designed to inform the debate regarding the existence of egocentric and allocentric representations of the environment in long-term memory, but the terms themselves lack the precision to be useful in the present discussion. Instead, a distinction was made between concrete, visual, snapshot-like representations and amodal, spatial representations composed of objects' locations and inter-object distances and directions. Visual representations are assumed to allow for quick judgments as to whether a stimulus has been previously experienced, using visual matching. They are presumed to be retinotopic. For example, a picture of a display from a view considerably closer to the display (but comprising all objects) would be expected to be recognized more slowly than a picture from the original view (Waller, 2006). Visual representations are necessarily egocentric, in that they are formed from a specific, experienced, point of view. It is assumed that these representations are used to determine one's present location within a familiar environment.

Spatial representations are assumed to contain object locations and inter-object distances and directions. Direct experience with the objects (i.e. viewing them) is not necessary to form a representation of their relative locations. For example, Avraamides, Loomis, Klatzky, and Golledge (2004) demonstrated equivalent performance of allocentric judgments when the spatial information was acquired from viewing a display or reading a description of a display. If spatial information is acquired through a sensory modality at learning, it is assumed that the sensory information is not maintained within the long-term representation. The features of individual objects are not contained within this representation. Spatial representations are not necessarily non-egocentric in nature[1]. Allocentric representations, however, are necessarily amodal, as they

[1] It is possible that, when an array of objects is experienced from a single position in the center of the array, individual self-to-object relations provide the most salient frame of reference by which to construct a long-term representation of the objects' locations. It is possible that in such circumstances the long-term representation of the objects' locations would be centered on the body, and that conversion to an allocentric representation (with a dominant axis congruent with the egocentric experience) would be necessary to perform object-to-object spatial judgments. The existence, or lack thereof, of egocentric abstract representations does not bear on the present discussion. The possibility is

lack an egocentric viewpoint, and contain only object-to-object spatial relations, organized along dominant axes and directions (see McNamara, 2003) or inter-object vectors (see Easton & Sholl, 1995; Sholl, 2000; Sholl & Nolin, 1997). It is impossible to encode the appearance of something, without encoding a viewpoint from which it was experienced. Spatial representations are proposed to support goal localization (Valiquette & McNamara, 2007).

No claim is being made that a task that is usually performed by accessing a visual representation cannot be solved by accessing a spatial representation, or that a task that is usually performed by accessing a spatial representation cannot be solved by accessing and manipulating a visual representation (if the spatial information has been visually experienced). Rather, it is being asserted that a cost is associated with relying on a spatial representation to perform a task that optimally relies on visual matching, or extrapolating from visual representations when allocentric information is necessary. It is also conceivable that a salient representation of object-to-object relations enhances recognition based primarily on visual matching and that visually presented cues aid in the recall of object-to-object relations. There is no precedence upon which to base predictions on the amount of influence of spatial information upon largely visual judgments and vise-versa. Therefore, speculation regarding the possible results of these influences was limited.

The following experiment was intended to determine if a single-type representational system such as that proposed by Mou et al. (2008) can

---

raised here merely to demonstrate that such a possibility is not inconsistent with the proposed distinctions between types of spatial memory.

accommodate behavior following exposure to more than a single view of an environment, or if a dual-type representational system is necessary to explain the full range of navigation-related behavior. A number of assumptions were made in order to distinguish between experimental results predicted by Mou et al's (2008) single-type representational system, which contains a single, amodal, spatial representation, and Valiquette and McNamara's (2007) dual-type representational system, which contains both an amodal, spatial representation and multiple visual representations.

Within the single-type system, viewing positions from which a configuration has been experienced are encoded with respect to the distance and direction from the configuration's intrinsic structure, within a single spatial representation. Based on that assertion, it was assumed that when multiple views are experienced they are represented with equal fidelity in the resulting long-term representation. This assumption conflicts with results of multiple experiments in which two or more views of a configuration were experienced and JRD showed a benefit for a single experienced view (Shelton & McNamara, 2001, 2004a, 2004b; Valiquette & McNamara, 2007; Valiquette et al., 2007).

One way to resolve this conflict is to assume that in previous JRD experiments the main effect of availability of intrinsic structure in the test stimuli overshadowed the main effect of familiarity of distance and direction to the intrinsic structure. Cues to the intrinsic structure of the layout and distance from the study view were independently manipulated in Mou et al.'s (2008) scene recognition experiments. In JRD these variables are not independent, which may account for why the separate effects found in Mou et al.'s study were not apparent in previous studies.

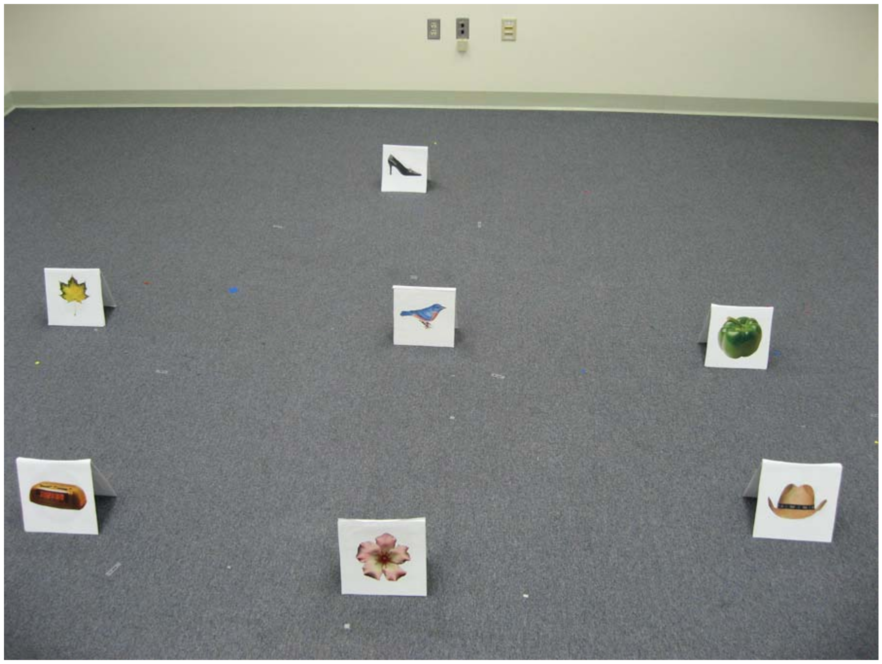
The two objects that define the heading in an aligned JRD trial necessarily provide a cue to the intrinsic structure. For example, the JRD trial "Imagine you are standing at the *book* facing the *wood*. Point to the *shoe,*" includes two objects that fall along the dominant 0° - 180° axis (see Figure 1). Cues to the intrinsic structure are absent in most JRD trials from misaligned headings (e.g., "Imagine you are standing at the *clock* facing the *shoe*. Point to the *banana.*") The two objects that define the position and facing direction are never aligned along an intrinsic axes, and, therefore do not provide a cue to the intrinsic structure. In some trials, however, the object that is to be pointed to and one of the other objects sometimes fall along an intrinsic axes (e.g., "Imagine you are standing at the *clock* facing the *shoe*. Point to the *scissors*."), which could be interpreted as providing a cue to the intrinsic structure. For this reason, a simple distinction cannot be made in which aligned JRD trials provide cues to the intrinsic structure of the layout, whereas misaligned trials do not.

It could be argued that JRD composed of pictures instead of words on a computer screen would enhance access to familiarity of distance and direction to the intrinsic structure, resulting in a benefit for the heading congruent with the misaligned study view over novel views. If this were the case, test headings congruent with the aligned study view should show a benefit over test headings congruent with the misaligned study view, due to the combined effects of familiarity and availability of cues to the intrinsic structure.

Alternatively, within the dual-type system, visual representations are formed at both viewing positions. Scene recognition is performed via visual matching to visual based representations. In addition, a single spatial representation of object-to-object relations is formed, the axes of which are congruent with axes within the layout that are made salient by, in the present case, their alignment with the walls of the study room. This representation is employed for goal localization (i.e. JRD). It is also possible that the visual representation within the dual-type system are accessed when JRD are presented visually. If this were the case, it would result in a benefit for test headings congruent with the misaligned study view over novel headings.

The JRD results generated by dual access to both one of the visual representations and the spatial representation would usually be indistinguishable from those predicted by access to Mou et al.'s (2008) single spatial representation. The experiment presented below, however, allows for these explanations to be distinguished.

In order to decouple the visual and the spatial information provided within a layout of objects, a new type of layout was constructed for this experiment. The objects comprising the layout looked markedly different depending upon the perspective from which they were viewed (0° or 135°; see Figures 3a & b). The seven objects that comprise the layout will be referred to as *objects*, whereas the pictures printed on the sides of the objects will be referred to as *images*. Each object contained two images – a 0° image and a 135° image. These images appeared on the 0° and 135° *sides* of the objects, respectively. Each object contained images that share a common name (e.g., *clock, flower, hat, leaf, bird, pepper, shoe*), but were visually distinct. The images were visible from either the 0° ($bird_0$) or the 135° ($bird_{135}$) viewing positions. This manipulation allowed for the spatial components of the layout to be pitted against the visual features of the layout in novel ways, using JRD, scene recognition, and priming.


a.


b.

Figure 3. Pictures of the layout (a) from the 0° study location, (b) from the 135° study location.

The priming task was designed to address a potential criticism of Valiquette and McNamara's (2007) proposal of two systems for human spatial memory, which is that scene recognition and JRD are very different in their performance demands. It could be a difference in performance demands, rather than a difference in representation accessed to perform the tasks, that produces the difference in the pattern of results (see Siegel, 1981).

One method for exploring relationships between components of a mental representation that has minimal performance demands is priming in item recognition. In studies of priming in item recognition (a type of associative priming), the response time to a stimulus (target) is analyzed with respect to the stimulus that preceded it (prime). If the response time to the target is faster when it is preceded by prime A than when it is preceded by prime B, the concept of prime A is assumed to be more closely associated than that of prime B to the concept of the target. This method has been used to investigate memory for sentences and paragraphs (McKoon & Ratcliff, 1980; Ratcliff & McKoon, 1978), the relationship between semantic properties of a concept to semantically and phonologically related concepts (McNamara & Healy, 1988), and the structure of mental representations of spatial relations (McNamara, Halpin, & Hardy, 1992; McNamara, Hardy, & Hirtle, 1989; McNamara, Ratcliff, & McKoon, 1984). McNamara and colleagues did not distinguish between the effects of spatial context and visual context on the mental representation of object locations.

The goal of the present priming task was to determine if evidence exists for associations among objects that are based on visual context, and thus for visual representations. In order to achieve this goal, relative response times to a target item were compared based on the item that directly preceded the target (the prime). Each trial began with the presentation of the prime, followed by a brief inter-stimulus-interval, and then by the target. When the target was presented, the participant decided whether it was part of the study layout (target trial) or not (distractor trial).

The relative strengths of spatial and visual representations have not previously been considered in studies of spatial memory or spatial priming. It is reasonable to assume, however, that the representation that dominates in a given situation is the one that is best suited to the task. The priming task used here was visually oriented. Therefore, the visual snapshot-like representations were expected dominate within the dual-type system, if they do indeed exist.

Within the framework of the present discussion concepts are *objects* within a spatial representation and *images* within visual representations. If connections exist between images, reaction time should be faster for an image if it is preceded by a nearby image that was perceived from the same view (bird$_0$ - shoe$_0$) than if it is preceded by an image located near in space but perceived from a different view (bird$_{135}$ - shoe$_0$). The images seen from the same viewing location and near each other are close in spatial and visual context, while the images seen from different views share only spatial context. Greater priming by a nearby image, that shares visual and spatial context, than by an image of a nearby object, that shares only spatial context, would provide evidence for visual representations, and thus a dual-type representational system.

It is assumed that the single-type representational system does not include the visual features of the objects. The only component of the objects that is represented is their locations with respect to each other. Because the spatial representation does not preserve visual features, visual context is lost. Therefore, $bird_0$ and $bird_{135}$ do not exist as separate entities within the spatial representation, and no difference in priming would be expected between images that share spatial and visual context ($bird_0$ - $shoe_0$), and those that share only spatial context ($bird_{135}$ - $shoe_0$).

The dual-type representational system predicts better performance for same-side trials (SS; see Figure 4, cells A & C), which shared visual as well as spatial context than for different-side trials (DS; cells B & D), which shared only spatial context. Within the dual-type system images (e.g. $bird_0$, $bird_{135}$) are encoded in the visual representations and therefore affect performance.

Target Trial Object / Name	Prime Type			
	Layout Image		Non-Layout Image	
	Same Side	Different Side	Familiar	Novel
Same	A $bird_0$ -> $bird_0$ mrt = 681(188) mdiff = 100	B $bird_{135}$ -> $bird_0$ mrt = 816(180) mdiff = -35	E $bird_f$ -> $bird_0$ mrt = 782(232)	F $bird_n$ -> $bird_0$ mrt = 774(192)
Different	C $shoe_0$ -> $bird_0$ mrt = 683(150) mdiff = 128	D $shoe_{135}$ -> $bird_0$ mrt = 765(179) mdiff = 45	G $shoe_f$ -> $bird_0$ mrt = 810(189)	H $shoe_n$ -> $bird_0$ mrt = 733(197)

$bird_0$, $shoe_0$ – images from layout seen from 0° viewing position
$bird_{135}$, $shoe_{135}$ – images from layout seen from 135° viewing position
$bird_f$, $shoe_f$ – images from memory game $bird_n$, $shoe_n$ – novel images
mrt – mean of participants' harmonic mean response times (standard deviation)
mdiff – mean of participants' difference scores

Figure 4. Priming experiment target trial types, response times and difference scores. ms = milliseconds.

The single-type representational system predicts equal amounts of priming on DS (see Figure 4, cells B & D) and SS (cells A & C). This pattern of results is predicted because visual representations, in which images are encoded, are not present in the single-type system, and therefore cannot influence performance.

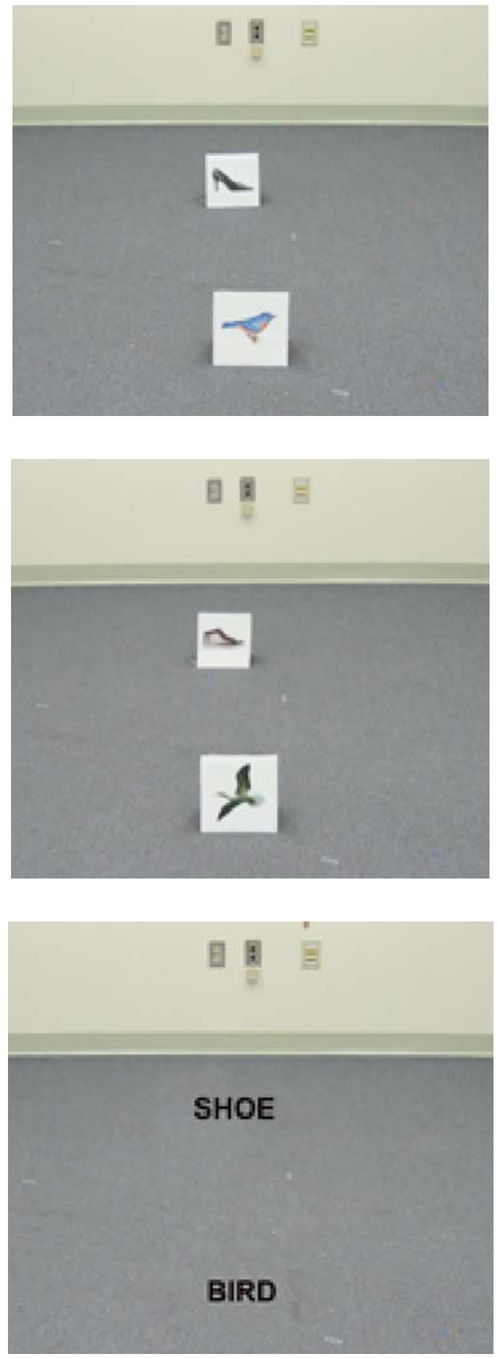
The stimuli used in JRD were modified for the present experiment. Heading information was provided pictorially rather than as verbal instructions on the computer screen. This manipulation helped to isolate the goal localization component of JRD by providing the positional information that must be imagined in standard JRD. Pictorially presented heading information also allowed for comparisons to be made based on visual differences among trials. For example, a picture that contains $bird_0$ and $shoe_0$ in their correct locations and a picture from the same location that contains $bird_{135}$ and $shoe_{135}$ in their correct locations provide the same spatial information, but different visual information. Lastly, visually presented heading information increases the similarity between JRD and scene recognition stimuli, decreasing the likelihood that differences in results between the two tasks are due to differences in the stimuli used.[2]

The difference between the patterns of performance predicted by the single-and dual-type systems in JRD is at the pictured heading corresponding to the study view misaligned with the dominant frame of reference (135°). In the dual-type representational system, the representation used to perform JRD does not maintain information about viewing positions that are not aligned with salient external frames of reference following exposure to a viewing position that is so aligned. Therefore, performance at the pictured heading of 135° is predicted to be no better than performance at novel pictured headings.

It is possible, however, that the visual representations within the dual-type system have a small effect on JRD results. Trials in which both a visual

[2] A pilot study, with 13 participants, was conducted to determine whether seeing a photograph of the heading to be imagined would produce substantially different results from those produced when heading information is presented as text. The patterns of error were very similar for standard JRD, consisting of words on a computer screen (e.g., "Imagine you are standing at the ___, facing the ____, point to the ____,"), and JRD in which the standing and facing objects were presented pictorially and pointing instructions were presented aurally.

representation and the spatial representation are accessed (i.e. trials from the 0° heading in which the 0° sides of objects are pictured; Figure 5a) could result in somewhat better performance than trials in which only the spatial representation is accessed (i.e. trials from the 0° heading in which the 135° sides of objects or object names are pictured; Figures 5b and 5c respectively). Trials in which only the visual representation is accessed (i.e. trials from the 135° heading in which the 135° sides of objects are pictured) might also show some benefit over other trials from the 135° heading.



a.

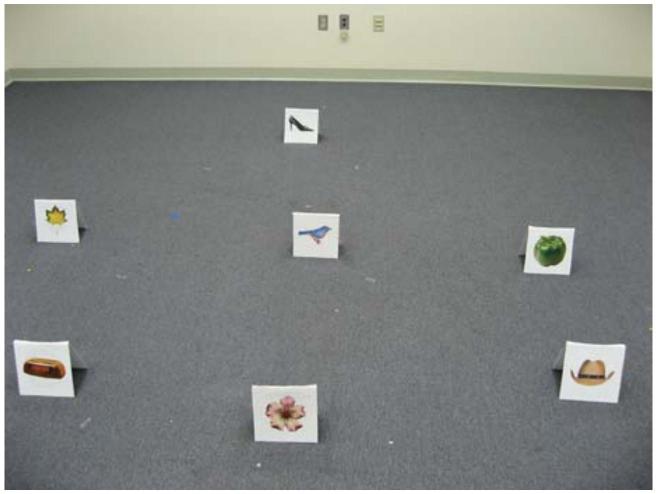
b.

c.

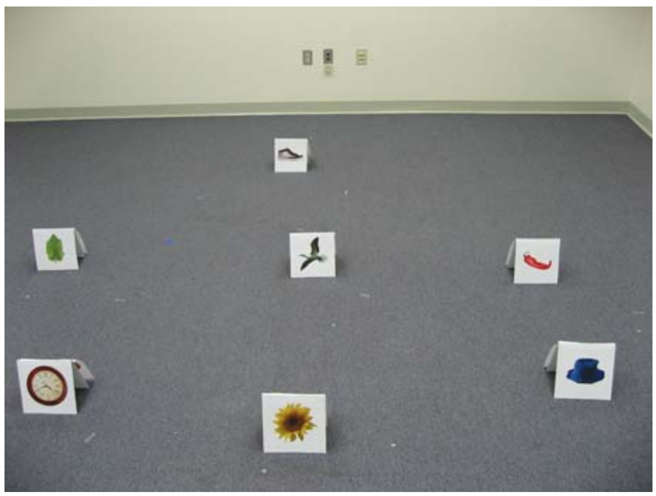
Figure 5. Examples of visual judgments of relative direction stimuli (a) 0° heading, 0° sides, (b) 0° heading, 135° sides, (c) 0° heading, text.

Alternatively, the pattern of results predicted by the single-type representational system shows a benefit for the misaligned experienced heading over novel misaligned headings. This benefit corresponds to the benefit for the experienced (but misaligned with the intrinsic structure of the layout) view found in Mou et al. 's (2008) experiments. The 0° pictured heading is predicted to show a benefit over both novel pictured headings and the experienced 135° heading because cues to the intrinsic structure of the layout are available in every test trial corresponding to the 0° heading (i.e. the objects are aligned with the intrinsic structure of the layout on every trial). The benefit for the aligned experienced heading over the misaligned experienced heading corresponds to the benefit for intrinsic triplets over non-intrinsic triplets in Mou et al.'s study (see Figure 2b). Some benefit is also predicted for test headings that are aligned with the salient orthogonal axes of the layout (90°, 180°, and 270°) because these trials will necessarily contain two objects, or words depicting objects' locations, that are aligned with the intrinsic structure of the layout on every trial (i.e. every trial will include two objects that fall along the 0° - 180° axis or the 90° - 270° axis).

The scene recognition task used in the present experiment differs little from the scene recognition task used by Valiquette and McNamara (2007). One group of eight target pictures was used in the previous experiment, whereas in the current experiment three groups of eight target pictures are used – one showing the sides of the objects viewed from the study position aligned with the walls of the room (Figure 6a), one showing the sides of the objects visible from the misaligned study position (Figure 6b), and one in which words replaced the objects (Figure 6c).



a.



b.

c.

Figure 6. Examples of scene recognition stimuli (a) target, 0° heading, 0° sides, (b) target, 0° heading, 135° sides, (c) target 0° heading, text.

The dual-type representational system contains multiple visual, snaphot-like representations. Therefore, representations from each viewing position should be housed in this system. The representations formed at the study positions should be most accessible when a matching picture of the layout is presented (i.e. one in which the 0° sides of the objects are shown from the 0° viewing perspective, or the 135° sides are shown from the 135° viewing perspective). This would result in performance that is best for pictures of the 0° sides and 135° sides from their respective study views, and decreases with angular distance from the study view. Furthermore, familiarity with the overall shape of the layout from the study views could generate partial matches between viewing perspective and object arrangement (i.e. the 135° images viewed from 0° as seen in Figure 6b) that result in some savings at the corresponding pictured headings. The trials involving the text layout (see Figure 6c) should be the most difficult of the three trial types if scene recognition is performed mainly through visual matching. A costly cognitive process of translating the spatial words into the appropriate visual images is predicted to be involved. This process could benefit some when the shape of the layout is familiar (at study views).

According to the single-type representational system, performance is expected to be better at both studied views than at novel views. This benefit is not due to familiarity with the visual content of the stimuli, but rather to matching between (a) the distance and direction from the viewing position to the intrinsic structure of the layout as it is portrayed in the picture and (b) the distance and direction from the viewing position to the intrinsic structure of the layout as it is represented in memory. When the distance and direction in the picture correspond to the distance and direction from a study position within the remembered representation, retrieval of inter-object spatial relations (from the viewpoint corresponding to the picture) is fast and accurate. When the distance and direction to the intrinsic structure of the layout as it is portrayed in the picture (i.e. the viewpoint of the picture) does not match the distance and direction from the viewer to the intrinsic structure of the layout as it is mentally represented, the inter-object spatial relations of the mental representation must somehow be manipulated (e.g., by mental rotation) before comparison can take place between the pictured configuration and the remembered configuration. This manipulation comes with a cost that appears in the results as an increase in reaction time for unfamiliar headings over the familiar studied headings.

CHAPTER II

EXPERIMENT

The experiment presented below was composed of five parts. Upon arrival participants took part in task that familiarized them with some of the images that were used in the priming task. The task was modeled after the childhood game referred to as “the memory game” or “concentration”. Upon completion of that task, they were taken to a different room for the learning phase. They then completed the priming, scene recognition, and JRD tasks (the order of which was counterbalanced) on a computer located in a third room.

The single- and dual-type representational systems predict different patterns of results for each of the three experimental tasks. In JRD, the single-type system predicts best performance for the aligned study heading and a benefit for the misaligned study heading over novel pictured headings. The dual-type system predicts best performance at the aligned study heading, with performance at the misaligned heading being no better than performance at novel, misaligned, pictured headings. It also predicts a modest benefit for study-heading trials in which the images in the picture match the images that were visible from that study position (see Figure 5a) over study-heading trials in which the images in the picture do not match the images that were visible from that study position (see Figure 5b). In scene recognition, the single-type representational system predicts no effect of the sides of the objects that are visible in the pictures on response time. The dual-type system predicts better performance at the pictured heading corresponding to the aligned study view (0°) when the images in the picture match the images that were visible from that view (0° images; see Figure 6a) than when the images in the picture match the images seen at the misaligned study view (135° images; see Figure 6b), and better performance at the 135° pictured heading for pictures of 135° images than for pictures of 0° images. The single-type representational system predicts no difference in priming between same-side and different-side prime-target pairs, whereas the dual-type system predicts a benefit for same-side pairs over different-side pairs.

Participants

Twenty-nine (15 female) undergraduates, graduate students, and members of the community (ages 18 – 42) participated in exchange for pay. Data from five participants (3 female) were removed – two because of high error on the scene recognition task, and three because of consistent error for target trials at one of the pictured headings in the scene recognition task, resulting in empty cells in the data analysis. Additionally, data from one participant were removed from the JRD analysis because that participant's data showed an extreme pattern that conflicted with the other participants' data and skewed the results (e.g. 57°, 62°, and 12° average error at the pictured headings of 0°, 180°, and 90°, respectively).

In the results of the priming task, trials in which an image of a given name (i.e. bird) appeared as the prime or target were removed for a given participant if he or she produced an error on more than twenty-five percent of the trials in which an image of that name was the target. Error of more than twenty-five percent was taken to indicate considerable uncertainty about where the individual images by that name had been encountered (in the study layout, or not). Six participants data were removed for trials containing images by one name (i.e. all trials containing “bird” images), three participants data were removed for trials containing images by two names (i.e. all trials containing “bird” images, and all trials containing “leaf” images).

Materials and Design

Familiarization Task

A memory game was constructed that included 14 pairs of images that fit within a 3 cm X 5 cm rectangle. Seven of the images shared names with the objects studied in the learning phase (e.g., *bird*, *shoe*), the other seven images, added solely to make the matching task more challenging, were common household objects (e.g., *teapot, banana*). The images were attached to a .6 m X .6 m board in a semi-random order. Cardboard covers were constructed that cover each of the images individually.

Learning Phase

A configuration of seven objects was used (see Figures 3a & b). The two sides of each object was 15 cm tall X 20 cm wide, and constructed of cardboard. Distinct images of the same name was attached to the sides of each object, producing a “clock” object, a “flower” object, a “hat” object, a “leaf” object, a “bird” object, a “pepper” object and a “shoe” object.

Judgments of Relative Direction.

Each JRD test trial was presented by computer and constructed from a picture of two objects, occupying their assigned locations within the study room, that depicted a specific heading (see Figures 5a, b, & c) and an auditory file containing pointing instructions. Participants were instructed to imagine standing at the object closest to the bottom of the picture and facing the other object. They were told to press a button to receive pointing instructions (e.g., “Point to the *bird*”) when they had assumed the to be imagined position. Pointing was performed with a joystick.

The primary independent variable was pictured heading. Headings were identified counterclockwise from 0° to 315° in 45° steps. Pointing direction, defined as the direction of the target object relative to the pictured heading, was varied systematically by dividing the space about a given heading into three regions: Front (0° - 60° & 300° - 360°), sides (60° - 120° & 240° - 300°), and back (120° - 240°). Pointing direction was counterbalanced across imagined headings. Participants received a total of 144 trials, 48 showing the 0° sides of the objects (see Figure 5a), 48 showing the 135° sides of the objects (see Figure 5b), and 48 in which text replaces the objects (see Figure 5c).

The principal dependent measure was the angular error of the pointing response, measured as the absolute angular difference between the judged pointing direction and the actual direction of the target. Two types of response latencies were also collected – orientation, measured from the appearance of the heading picture to the button press that initiated the pointing instructions, and pointing, measured from the offset of the pointing instruction to the completion of the pointing response.

Scene Recognition

A distractor layout was produced by reflecting the objects along the 135° - 315° axis, creating a mirror reversal of object locations along that axis[3]. Each scene recognition test trial required participants to determine if a picture presented on a computer screen was of the configuration of objects that they had studied, or a picture of the same objects occupying different locations (i.e., the distractor layout). The primary independent variable was pictured heading. Eight equally spaced headings were used. Headings was identified counterclockwise

[3] The axis of reflection was chosen because it was consistent with one of the axes of reflection used in a recent scene recognition experiment in our lab. In that experiment distractor layouts were constructed by reflecting along the 0° - 180° axis and the 135° - 315° axis. Participants studied a layout from the 0° and 135° viewing positions. The axis of reflection of the distractor layout was shown not to impact performance on the target trials.

from 0° to 315° in 45° steps. Participants received a total of 288 trials, composed of six blocks of 48 randomly presented trials, three trials at each target pictured heading (a picture of the studied layout) – one showing the 0° sides of the objects (see Figure 6a), one showing the 135° sides of the objects (see Figure 6b), and one in which text replaced the objects (see Figure 6c) - and three corresponding distractor trials, with a break between blocks two and three.

The principal dependent measure was response latency on trials in which the target layout was pictured. Error data was also collected.

Priming

The priming task consisted of 224 trials. These included 112 trials in which the target image was from the layout (target trials, see Figure 4). Fifty-six of these prime-target pairs contained prime and target images from the layout: 14 same object-same side trials (SOSS; cell A), 14 different object-same side trials (DOSS; cell B), 14 different object-same side trials (DOSS; cell C), 14 different object-different side trials (DODS; cell D). The other 56 prime-target pairs contained a non-layout prime image (either an image from the memory game with the same name as one of the layout objects, or a novel image with the same name as one of the layout objects). Of those 56 trials, 14 contained a same name-familiar image as the prime (SNF, cell E), 14 contained a same name-novel image as the prime (SNN, cell F), 14 contained a different name-familiar image as the prime (DNF, cell G), and 14 contained a different name-novel image as the prime (DNN, cell H). Each of the layout images was used once as the target in each of the eight target trial conditions, resulting in 14 trials per condition.

The remaining 112 trials were distractor trials in which the target images were not from the layout. Two sets of images that were not part of the study layout were used as targets in the distractor trials. The first set was composed of the seven images from the memory game that had the same names as the seven layout objects. The second set consisted of novel images with the same names as the layout objects.

Each image from the layout was presented 12 times throughout the course of the experiment. The trials were presented in a set order that balanced both trial type and image repetition.

Procedure

Familiarization Task

Participants played the memory game prior to engaging in the learning phase in order to become familiar with some of the images used as distractors in the priming task. They were instructed to uncover two images at a time. If the images matched, they did not replace the covers. If they did not match, they replaced the covers and tried again. The game was finished when every image was visible. When the participants finished the memory game, they were escorted to the study room for the learning phase.

Learning Phase

Order of viewing (0° -135°, 135° - 0°) was counterbalanced across gender. Before entering the study room, each participant was instructed to learn the objects and their locations for a spatial memory test. The participant was blindfolded and led into the study room to the first viewing position. The blindfold was removed, and the experimenter named the objects while pointing to them in a random order. The participant viewed the display for 30 seconds before being asked to close his or her eyes. The experimenter then named the objects in a random order, as the participant pointed to them. The first study phase was complete when the participant correctly pointed to each of the objects on two consecutive trials, and then, after studying the layout the last time, correctly named and pointed to each of the objects with his or her eyes closed. The participant was then blindfolded and led to the second viewing position, and the learning process was repeated. When the participant has successfully learned the objects' locations, the blindfold was replaced and the participant was led from the room.

Judgments of Relative Direction

Testing took place on a different floor of the building. The JRD test trials were presented on a Macintosh computer. Five practice trials were completed with the experimenter present. Feedback was given during practice to ensure that the participant understood the task. Participants pressed a button on the joystick to begin a trial. A heading picture appeared on the screen (see Figures

5a, b, and c). When the participant adopted the pictured position, he or she pressed a button on the joystick and received aurally presented pointing instructions (e.g. “ Point to the *bird*.”).

Scene Recognition

The scene recognition test trials were presented on a Macintosh computer, using SuperLab Pro 1.75. Participants were instructed to press the "J" key if the picture presented was of the layout that they had studied and to press the "F" key if it was not. The experimenter remained in the room for the first four trials to answer any questions. Trials performed while the experimenter was present were removed from the data prior to analysis.

Priming

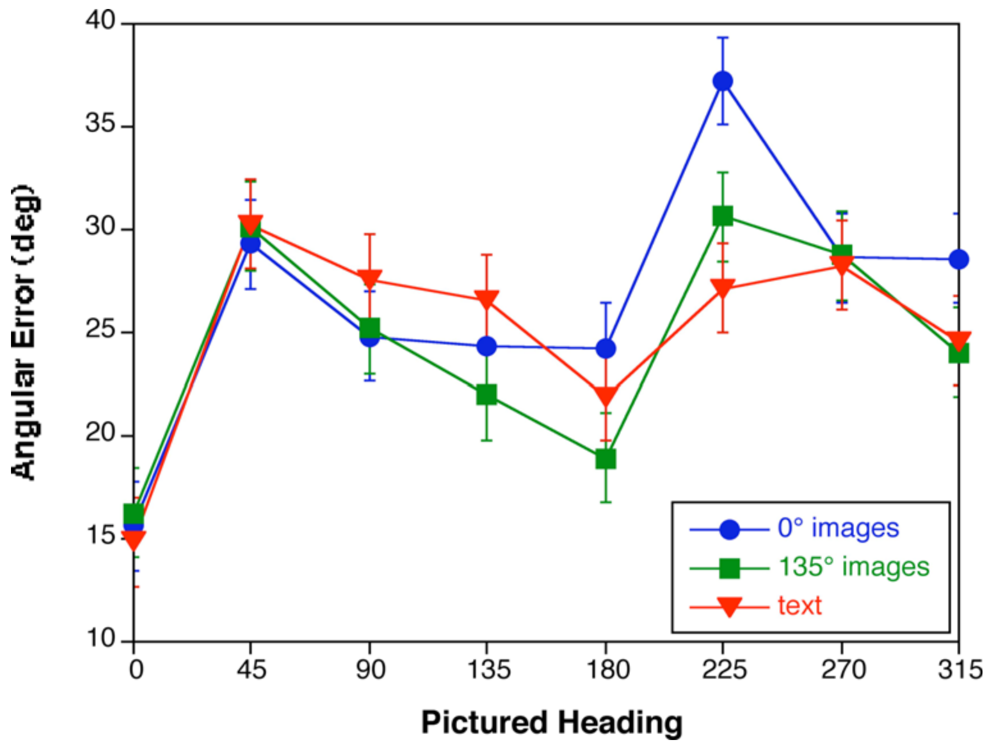
The priming trials were presented on a Macintosh computer, using SuperLab Pro 1.75. The participant pressed the space bar to begin a trial. When the spacebar was pressed, the prime image appeared on the computer screen for 250 ms. The participant was instructed not to respond to the prime image. After a 150 ms inter-stimulus interval, the target image appeared. The participant was instructed to press the “J” key to indicate that the image was part of the study layout, and press the “F” key to indicate that it was not. They were instructed to respond as quickly as possible. When a response was given, they were prompted to press the spacebar to begin a new trial.

Participants performed the priming, scene recognition, and JRD tasks in one of three orders – priming, scene recognition, JRD; scene recognition, JRD, priming; JRD, priming, scene recognition. Assignment to testing order was pseudo-random and balanced for study order and gender.

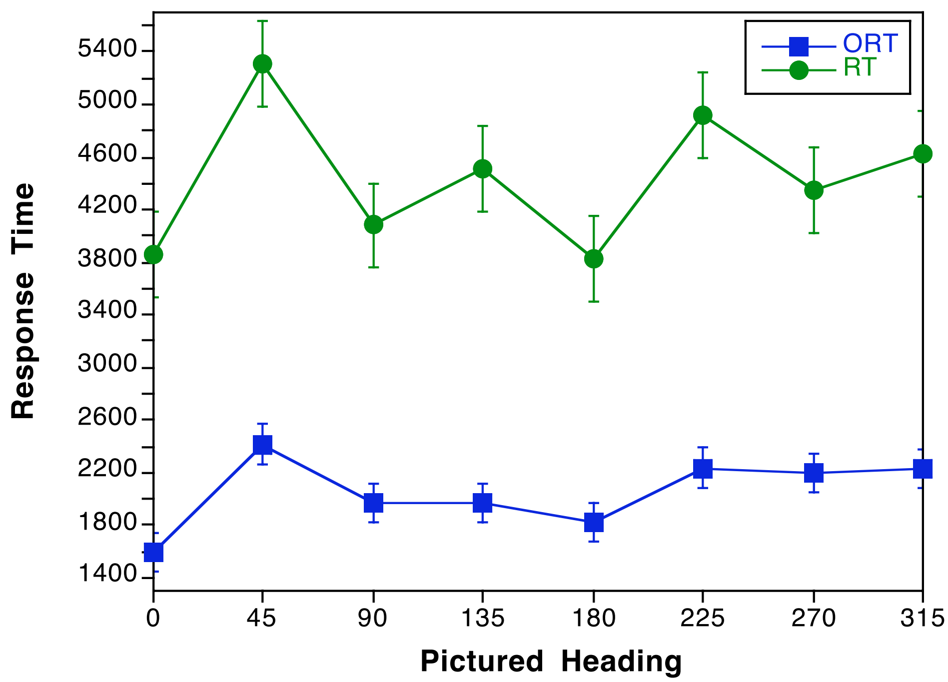
Results and Discussion

Judgments of Relative Direction

Mean absolute error, collapsed across participants, is presented in Figure 7a as a function of pictured heading and image type. The results show similar patterns across headings for each of the three image types, with better performance for the 0° pictured heading than for all other headings. Performance was only slightly better for the 135° pictured heading than for other misaligned headings.



a.



b.

Figure 7. Judgments of relative direction (a) angular error as a function of pictured heading (b) response time as a function of pictured heading. deg = degrees. ms = milliseconds.

Statistical analysis supports this conclusion. Mean absolute angular error was calculated for each participant and each condition, and analyzed in a split-plot ANOVA with terms for order of viewing (0° - 135° or 135° - 0°), gender, order of testing (scene recognition (SR) - JRD – priming, JRD – priming – SR, priming – SR – JRD), image type (0°, 135°, or text), and pictured heading (0° - 315°, in 45° increments). Image type and pictured heading were within participants.

The main effect of pictured heading was significant, $F\ (7,84) = 8.45$, $p < .001$, $MSE = 191.64$. A contrast comparing average error at the 135° pictured heading to average error at novel misaligned headings was not significant, F (1, 84) = 1.47. Paired t-tests revealed that performance was better for the 0° pictured heading than for all other pictured headings, *t*s (84) ≥ 2.66, $p < .05$, and that performance was no better for the studied 135° pictured heading than for novel pictured headings, *t*s (84) ≤ 1.3. Performance was better at the novel pictured heading of 180° than at four other novel pictured headings (45°, 90°, 225°, and 270°), *t*s (84) ≥ 2.26, $p < .05$. A contrast comparing performance for 135° images to performance for 0° images and text at the 135° pictured heading was not significant, $F\ (1, 84) < 1$, nor was contrast between 135° images and the others at the 180° pictured heading, $F\ (1, 84) = 1.08$. No other main effects or interactions were significant.

Two sets of response times were also analyzed, and are shown in Figure 7b. Orientation response time (ORT) is the time between the presentation of the pictured heading and the button press that initiates the aural presentation of the pointing instructions (e.g. “Point to the *bird*.”) Pointing response time (RT) is the time between the conclusion of the pointing instructions and the completion of pointing. Very little difference across pictured heading is apparent in ORT. Pointing response time results, however, show the saw-tooth pattern of faster responses for pictured headings aligned with the walls of the study room.

Statistical analyses support these conclusions. Harmonic means were calculated for each participant and each condition, and analyzed in two split-plot ANOVAs (ORT and RT), with terms for order of testing, study order, gender, image type, and pictured heading. The main effect of pictured heading was significant for ORT, F (7,84), $p < .01$, $MSE = 962062$. Responses were faster for the studied 0° pictured heading than for all other pictured headings, except 180°, ts (84) ≥ 2.12, $p < .05$. The ORT analysis revealed no other significant effects or interactions.

The main effect of pictured heading was significant in the RT analysis, F (7,84), $p < .05$, MSE = 4656587. Response times were faster for pictured headings aligned with the wall of the study room (0°, 90°, 180°, 270°) than for misaligned, pictured headings (45°, 135°, 225°, 315°), F (1,84), $p < .05$, MSE = 4656587. The RT analysis revealed no other significant main effects or interactions.

The present JRD results support the dual-type representational system. No benefit was found for the studied 135° pictured heading over novel pictured headings, which would be predicted by the single-type representational system. Furthermore, no interaction between image type and pictured heading was evident in the results, indicating that visual information is not preserved in the representation accessed to perform JRD.

A saw-tooth pattern, in which performance was better for pictured headings aligned with the walls of the study room than for misaligned-pictured headings, is common among JRD results, and consistent with both the single- and the dual-type representational systems. This pattern was not apparent in the error results, but did emerge in the RT results. A possible explanation for the absence of the saw-tooth patter in the error data is that the layout used in this experiment was bilaterally symmetrical along the 0° - 180° axis and positioned on the bare floor (as opposed to on a square mat) of a room, so that the 0° - 180° axis of the layout was parallel to the long walls of the room. It is possible that the salience of the 90° - 270° axis was diminished by the removal of the mat, and that the salience of the 0° - 180° axis was enhanced by the bilateral symmetry, resulting in the locations of the objects being represented primarily with respect to the more salient 0° - 180° axis.

Scene Recognition

Mean response time in the scene recognition task, collapsed across subjects, and with error trials removed, is plotted in Figure 8 as a function of image type and pictured heading. Response times were faster for the 0° pictured heading than for the 135° pictured heading when the 0° images were shown, and faster for the 135° pictured heading than for the 0° pictured heading when 135° images were shown.

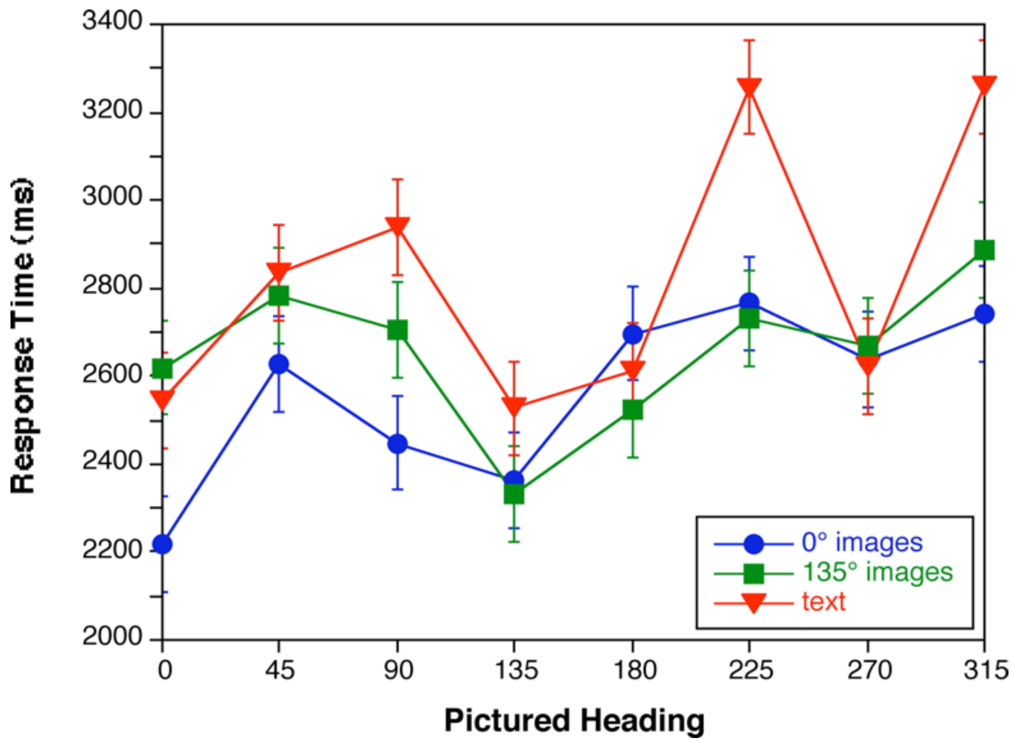


Figure 8. Response times in scene recognition as a function of pictured heading and image type. ms = milliseconds.

Statistical analyses support this conclusion. Error trials were removed (3.45 % of trials), and harmonic mean response times were computed for each participant in each condition. Harmonic mean response time was calculated for each participant and each condition, and analyzed in a split-plot ANOVA with terms for order of testing, study order, gender, image type, and heading. Order of testing, study order, and gender were between participants.

The interaction between image type and pictured heading was marginally significant, F (14, 168) = 1.72, p = .056, MSE = 280316. The interaction contrast between 0° and 135° images at the pictured headings of 0° and 135° was significant, F (1, 168) = 8, p < .01, MSE = 280316. A second contrast revealed a significant difference in performance between the aligned pictured headings of

180° and 270°, and the misaligned pictured headings of 225° and 315° for the pictures in which text replaced the images, F (1, 168) = 17.57, $p < .01$, MSE = 280316.

The main effect of image type was significant F (2, 24) = 5.11, $p < .05$, MSE = 669407. Response times were faster for pictures composed of the 0° or 135° images than for pictures in which text replaced the images, with mean response times of 2562 ms, 2656 ms, and 2824 ms, respectively. A significant interaction between image type and gender was also revealed, F (2, 24) = 4.13, $p < .05$, MSE = 669407, which was driven by the male participants whose response times were fastest for pictures containing 0° images (2485 ms), followed by 135° images (2613 ms) and text (2979 ms). Female participants produced nearly identical response times for the three image types (2639, 2699, and 2669 for 0° images, 135° images, and text, respectively). The main effect of image type was subsumed by a three-way interaction among image type, order of testing, and study order, F (2,24) = 5.54, $p < .01$, MSE = 669407. There were only data from four participants per cell in this interaction, and it did not admit to straightforward interpretation.

The main effect of pictured heading was also significant, F (1, 84) = 9.28, $p < .001$, MSE = 300122. There was no difference in response times between the 0° and 135° pictured headings, $t < 1$. Response times were faster for the pictured headings of 0° and 135° than for novel pictured headings ts (84) ≥ 2.1, $p < .05$. The interaction between pictured heading and order of testing was also significant, F (1, 84) = 2.09, $p < .05$, MSE = 300122. Participants who completed the scene recognition test first produced longer response times (3421 ms) than those produced by participants who completed the priming test (2296 ms), or JRD and priming (2447 ms), before scene recognition. Participants who completed the scene recognition task first also showed greater savings at the experienced pictured headings of 0° and 135°, compared to novel pictured headings, (532 ms, 15%) than the savings of participants in the other groups (210 ms, 9%). No other main effects or interactions reached significance.

The interaction between the 0° and 135° image types at the 0° and 135° pictured headings supports the dual-type representational system, which contains snapshot-like visual images of experienced views of the environment, over the single-type representational system, which does not store visual information. The picture of the 0° images taken from the familiar 0° viewing perspective was recognized more quickly than the picture of the 0° images taken from the familiar 135° perspective, and vise-versa. The benefit for pictures in which the side of the object (image) presented at one of the familiar pictured heading matched the side of the object that was visible from that study position (see Figure 6a), over pictures in which the visible side of the object did not match what was visible from the pictured study position (see Figure 6b), indicates that the objects' visual properties were encoded in the representation that was accessed to make the recognition judgment.

The main effect of heading, with savings for the 0° and 135° familiar pictured headings is consistent with the single-type representational system. The savings for the 0° and 135° familiar pictured heading can be attributed to familiarity with the distance and direction from the pictured viewing position to the intrinsic structure of the layout. This effect, however, was subsumed by the interaction between image type and pictured heading.

The saw-tooth pattern seen at the pictured headings of 180° - 315° for the pictures in which text replaced the images was unexpected. This pattern could indicate that participants accessed their spatial representations of the layout to determine if the words were in the correct configuration when neither the shape of the layout nor the images comprising the layout matched their visual representations. The results for the 90° pictured heading contradict that interpretation, however.

The main effect of image type and the interaction between image type and gender were not anticipated. The slower response times for males when the images in the pictures were replaced with the names of the objects (see Figure 8c) is consistent with the female preference and advantage for verbal processing that has been demonstrated repeatedly, across a wide range of tasks (Halpern, 2000). This finding does not inform the present discussion and will not be discussed further.

Priming

The response times for the target trials in the priming task are presented in Figure 4. Response times were faster for same side trials (cells A & C) than for different side trials (cells B & D).

Statistical analyses supported these conclusions. Error trials and distractor trials were removed. Harmonic means were generated for every participant and every trial type. Because of the nested nature of the variables (i.e. layout primes were on the same of the objects or on different sides, whereas non-layout primes differed in familiarity) the data for all target trials was not analyzed within a single ANOVA. Instead, a split plot ANOVA was performed on the data from the trials in which images from the layout were used as primes, and included terms for study order, testing order, gender, semantic relationship (whether prime and target share a name), and sides of objects presented (same - both 0° images or 135° images or different - one 0° and one 135° image). Study order, testing order, and gender were between subjects. The ANOVA was performed in order to generate an error term for the planned paired comparisons of difference scores discussed below, and to test for effects of between-subject variables and interactions among the between- and within-subject variables.

Individuals' response times in the conditions of interest were subtracted from their different-name familiar-prime trial (SNF; cell E in Figure 4) response time to generate a difference measure of response time. Individual participants' average response times on SNF served as the baseline for same-object same-side trials (SOSS; cell A) and same-object different-side trials (SODS; cell B). Individual participants' average response times on different-name familiar-prime trials (DNF; cell G) served as the baseline response time for different-object same-side trials (DOSS; cell C) and different-object different-side trials (DODS; cell D). Using a different baseline for different trial types isolated the component of interest within each trial type to an equal degree. In SOSS and SODS the prime and target share a name, whereas in the other two trial types of interest they do not. In order to remove the effect of shared name from the response times of SOSS and SODS, the response time baseline was from a trial type in which the prime and target also share a name (SNF). Response time difference was the dependent variable of interest.

The comparison between SOSS (cell A) and SODS (cell B) was significant, t (1, 12) = 7, $p < .05$, MSE = 4504, as was the comparison between DOSS (cell C) and DODS (cell D), t (1, 12) = 4.25, $p < .05$, MSE = 4504. The primes and targets in same-side trials share visual as well as spatial context (the images were seen from the same viewing position and close together in space), whereas the primes and targets in different-side trials share only spatial context (they were seen from different views but are close together in space).[4]

The comparison between SODS (cell B) trials and DOSS (cell C) was also significant, t (1, 12) = 8.4, $p < .01$, MSE = 4504, as was the comparison between SODS (cell B) and DODS (cell D), t (1, 12) = 4.14, $p < .05$, MSE = 4504.

The ANOVA revealed a significant main effect of sides of objects presented, F (1, 12) = 33.50, $p < .01$, MSE = 8239, and a significant interaction between sides of objects presented and study order, F (1, 12) = 5.55, $p < .05$, MSE = 8239. The interaction between sides of objects presented and semantic relationship was not significant, nor were any other main effects or interactions.

[4] In SOSS (cell A), the same image serves as both prime and target. Repetition priming, which is categorically different than associative priming, was expected to occur on this type of trial (McNamara, 2005). It cannot be determined the extent to which repetition priming affected response times in this condition.

The main effect of sides of object and the significant differences between the same-side prime and different-side prime trials for both levels of semantic relationship strongly demonstrate an effect of visual context on priming. This finding supports the dual-type representational system, which contains both visual and spatial representations of the environment, over the single-type representational system, which contains only spatial representations.

The significant difference between difference scores for SODS (cell B) and DOSS (cell C) could indicate that visual context overrode spatial context in this priming task. Images that occupied virtually the same space but differed in visual context (i.e., different sides of one object; cell B) produced less priming than FSN (cell E), whereas images that occupied different, but nearby spaces, and shared visual context (DOSS; cell C) produced more priming than FDN. This interpretation is compromised, however, by the direction of the significant difference between SODS (cell B) and DODS (cell D) difference scores. Response times for SODS (cell B) were slower than response times for corresponding familiar-prime trials, whereas response times for DODS (cell D) trials were faster than response times for corresponding familiar-prime trials, which would indicate a cost for semantic relatedness (sharing a name). Such a cost is highly counterintuitive. Therefore, no claim will be made that visual context trumps spatial context based on the comparisons of SODS (cell B) and DOSS (cell C) difference scores.

The interaction between sides of objects presented and study order was unanticipated. Participants who studied from the 135° position first responded more quickly than those who studied from the 0° position first when the prime and target were on the same side(s) of the object(s) (658 ms and 705 ms, respectively). The reverse was true when the prime and target were on different sides (810 ms and 770, respectively). No explanation for this between-groups difference is apparent at this time.

A second split plot ANOVA was performed on the response time data from the target trials that included familiar or novel primes (cells E – H in Figure 11). It included terms for study order, testing order, gender, semantic relation (same vs. different name) and familiarity (familiar vs. novel). It revealed a main effect of familiarity, $F(1, 12) = 8.46$, $p < .05$, $MSE = 4817$. Participants responded more quickly to target images when they were preceded by novel images than when they were preceded by familiar images from the memory game.

The difference between familiar- and novel-prime trial response times raises the possibility that conflicting visual and spatial context between prime and target slows response times. The familiar primes were experienced within the visual and spatial context of the memory game, whereas the target images were experienced within the visual and spatial context of the study room. Primes from a familiar, but conflicting, visual and spatial context could have interfered with processing of the target images, whereas neutral primes, which were devoid of visuo-spatial context, could not have. Further research is necessary to determine if indeed conflicting visual or spatial context between prime and target produces interference.

Mean response times for the distractor trial types were computed, and were between 690 ms and 775 ms. They were not submitted to further analysis because different criteria could have been used for each type of target. Familiar targets (from the memory game) required a decision regarding where the image was seen (upstairs or not), whereas novel targets (that were not seen prior to testing) could have been responded to based on familiarity (old or new).

The results of the priming task support the dual-type representational system, which contains both visual and spatial representations, over the single-type system, which contains only spatial representations. They show that when decisions must be made about stimuli that are visually presented, shared visual context between prime and target results in greater priming than shared spatial context alone.

Two differences that were expected, but were not apparent in the results, also bear noting. There was no measurable benefit for trials containing semantically-related (same-name) primes compared to trials containing semantically-unrelated primes, nor was there evidence of repetition priming for the same name-same side trials. It is not within the scope of the present discussion to speculate as to why these effects did not occur.

CHAPTER III

GENERAL DISCUSSION

The experiment presented here distinguished between two types of long-term spatial memory representational systems. The single-type representational system, proposed by Mou et al. (2008), and the dual-type representational system, proposed by Valiquette and McNamara (2007), predict distinct patterns of results for each of the component tasks. Mou et al.'s system is composed of a single, spatial representation of object-to-object relations. The representation is organized with respect to the dominant intrinsic axes or directions within the array of objects. Distance and direction from the viewing position to the intrinsic structure of the array are also encoded within the same representation. Valiquette and McNamara's representational system contains a spatial representation similar to that proposed by Mou et al., but without encoded viewing distance and direction information, as well as multiple snapshot-like visual representations.

Results of the JRD, scene recognition, and priming tasks in this experiment support Valiquette and McNamara's (2007) dual-type system. In the JRD results, the single-type system predicts a benefit for the pictured heading of 135° over novel pictured headings whereas the dual-type system predicts that performance at the 135° pictured heading would not differ from that at novel misaligned pictured headings. Neither system predicts differences in performance based on visual differences among stimuli (i.e. pictured heading depicted by the 0° sides of objects vs. pictured headings depicted by the 135° sides of objects). The results showed no benefit for the 135° pictured heading over novel misaligned headings.

In scene recognition, the dual-type system correctly predicts that the patterns of results produced would depend on the visual information available (i.e. visible images of objects or text) in the heading pictures: 0° images were recognized more quickly from the 0° heading than from the 135° heading, and 135° images were recognized more quickly from the 135° heading than from the 0° heading. The single-type system predicts that the patterns of results in scene recognition would not differ as a function of the visual information provided in the heading pictures.

Finally, in the priming task, the dual-type system predicts that pairs of images that share a common side (e.g., $bird_0$, $leaf_0$) would show greater priming than pairs of images that do not (e.g., $bird_0$, $leaf_{135}$). The single-type system predicts equal amounts of priming between pairs of images that shared a common side and those that do not. Results were, again, consistent with the dual-type system.

The results of the current study strongly indicate that visual representations of the environment exist in long-term memory, and that these representations affect recognition judgments but not judgments of the relative locations of objects. These findings support Valiquette and McNamara's (2007) assertion that multiple snapshot-like representations are accessed to determine one's location, while a separate allocentric representation is accessed to determine the location of an unseen goal.

The above claim that the dual-type system predicts the results of the present experiment, whereas the single type system does not, requires some qualification. A number of assumptions were made in order to make the predictions based on the single-type system. One assumption was that multiple views of the environment are represented with equal fidelity in the single representation within the system described by Mou et al. (2008). This assumption established the difference in JRD results predicted by the competing systems. It is difficult to reconcile this assumption with previous JRD results. It is possible that the distances and directions from multiple study views to the intrinsic structure of the layout are not encoded in the kind of representation proposed by Mou et al.. It is possible that, under the present study conditions, they would claim that only the most salient view would be represented long-term. That assertion, however, conflicts with the present, and previous (Valiquette & McNamara, 2007), scene recognition results. The other assumption, which established the predicted differences in scene recognition and priming results, is less controversial. It is that visual information is not accessible by scene recognition and priming because it is not represented within Mou et al.'s single representation. Mou et al. clearly stated that the long-term spatial representation that they proposed did not include visual information. It is possible, however, that visual information housed within some other representation could be accessed to make judgments about stimuli that differ in visual, but not spatial properties.

Regardless of the accuracy of the assumptions made about Mou et al.'s proposed representation, the present results provide strong evidence that visual representations of the environment, which facilitate place recognition, are housed in long-term memory.

The present experiment is the first to use objects with visually distinct sides to decouple visual and spatial long-term representations. This type of object could be used to further explore the roles of visual and spatial representations in human navigation. For example, it is possible that the amount of experience a person has with a given environment impacts the relative salience of spatial and visual representations of that environment. Visual representations might be formed quickly, whereas spatial representations are formed with extended exposure (Valiquette et al., 2007). It is possible that very limited exposure to both study views would lead to an increase in dependence on visual representations across testing conditions. This could lead to JRD results that have a pattern similar to that previously seen in scene recognition, with response times increasing as an effect of rotational distance from the study views. The effect of reduced study time on priming is more difficult to predict. It is possible that interference from conflicting visual context would increase, resulting in response times for different-side trials that are slower than those presented here.

Placing time limits on responses in the scene recognition and JRD tasks could be informative as well. The priming results indicate that visual information might be more easily accessed in memory than spatial information. Therefore, it is possible that visual representations might be accessed to perform JRD, which would lead to a pattern in the JRD results that resembles the pattern seen in scene recognition. If time limits were placed on responding in the scene recognition task, error would be expected to increase dramatically, except for the pictures that accurately portray the layout as it appeared from the two study views (0° sides of objects at 0° heading – 135° sides of objects at 135° heading).

The priming task was the first to attempt to separate the effects of visual and spatial context. This could be a fruitful avenue to explore.   However, it appears that the semantic relationship between the images on the sides of the objects produced unexpected interference in this experiment, which appeared as less priming between same-name prime-target pairs than between different-name prime-target pairs. This interference could have arisen from participants attending to the differences among the same-name images, rather than interpreting the 0° and 135° images as two views of the same whole. Interference might be reduced by using objects that customarily have distinct sides. A layout in which the sides of the objects depict the fronts and backs of unique houses (i.e. Tim's house, Adrianne's house, etc.) is one example. This would facilitate participants representing the objects in a holistic manner. Un-named pictures of fronts and backs of houses could then be used as familiar and novel primes and targets.

As informative as the present experiment might be, no claim is being made about the extent to which the results obtained in a highly controlled and contrived laboratory experiment can inform us about what happens in the real world. The participants in this study learned the locations of the objects because they were told to do so, not because learning the objects' locations would benefit them in some way (i.e. allow them to navigate through a new environment). The objects had no inherent meaning, and were specifically chosen for their limited semantic relatedness to each other. Objects in the real world differ in their degree of relatedness to each other, which can affect how their relative locations are represented in memory (see McNamara et al., 1989). Participants could see every object from both viewing positions, which were on the periphery of the layout, whereas we usually learn the locations of objects within our environment from a position within the environment where some objects obstruct the views of other objects. In other words, it is not clear to what degree the representations of the study layout formed by the participants in this study resemble the representations of his or her classroom formed by the student in the example given earlier.

In conclusion, the experiment presented here demonstrated a new approach to resolving whether both spatial and visual representations are needed to explain human navigation. The experiment was intended to clarify whether recognizing a place and locating an unseen goal access the same or different types of information. The results strongly suggest that recognizing a place depends on visual snapshot-like long-term representations, whereas locating an unseen goal accesses a single amodal, spatial long-term representation.

REFERENCES

Avraamides, M. N., Loomis, J. M., Klatzky, R. L., & Golledge, R. G. (2004). Functional Equivalence of Spatial Representations Derived From Vision and Language: Evidence From Allocentric Judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 801-814.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review, 114*(2), 340-375.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*(4), 302-307.

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*(2), 483-500.

Halpern, D. (2000). *Sex Differences in Cognitive Abilities*. Mahwah, NJ: Lawrence Erlbaum Associates.

McKoon, G., & Ratcliff, R. (1980). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 369-386.

McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In C. Freksa, W. Brauer, C. Habel & K. F. Wender (Eds.), *Spatial cognition III: Routes and navigation, human memory and learning, spatial representation and spatial reasoning, LNAI 2685* (pp. 174-191). Berlin: Springer.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Semantic priming: Perspectives from memory and word recognition New York, NY, US: Psychology Press.

McNamara, T. P., Halpin, J. A., & Hardy, J. K. (1992). Spatial and temporal contributions to the structure of spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 555-564.

McNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 211-227.

McNamara, T. P., & Healy, A. F. (1988). Semantic, phonological, and mediated priming in reading and lexical decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 398-409.

McNamara, T. P., Ratcliff, R., & McKoon, G. (1984). The mental representation of knowledge acquired from maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 723-732.

McNamara, T. P., Rump, B., & Werner, S. (2003). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonomic Bulletin and Review, 10*(3), 589-595.

Mou, W., Fan, Y., McNamara, T. P., & Owen, C. B. (2008). Intrinsic Frames of Reference and Egocentric Viewpoints in Scene Recognition. *Cognition, 106*(2), 750-769.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 162-170.

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and Egocentric Updating of Spatial Memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 142-157.

Mou, W., Zhang, K., & McNamara, T. P. (2004). Frames of Reference in Spatial Memories Acquired From Language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 171-180.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17*(4), 403-417.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. *Cognitive Psychology, 43*(4), 274-310.

Shelton, A. L., & McNamara, T. P. (2004a). Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*(1), 158-170.

Shelton, A. L., & McNamara, T. P. (2004b). Spatial memory and perspective taking. *Memory & Cognition, 32*(3), 416-426.

Sholl, M. J. (2000). The functional separability of self-reference and object-to-object systems in spatial memory. In S. O. Nuallain (Ed.), *Spatial cognition: Foundations and applications: Selected papers from Mind III,* *annual conference of the Cognitive Science Society of Ireland, 1998* (pp. 45-67). Amsterdam: John Benjamins.

Sholl, M. J. (2001). The role of a self-reference system in spatial navigation. In D. Montello (Ed.), *Proceedings of the COSIT International Conference: Vol. 2205. Spatial information theory: Foundations of geographic information science* (pp. 217-232). Berlin: Springer.

Sholl, M. J., & Nolin, T. L. (1997). Orientation specificity in representations of place. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(6), 1494-1507.

Siegel, A. W. (1981). The externalization of cognitive maps by children and adults: In search of ways to ask better questions. In L. S. Liben, A. H. Patterson & N. Newcombe (Eds.), *Spatial representations and ehavior across the life span.* New York: Academic Press.

Valiquette, C. M., & McNamara, T. P. (2007). Different mental representations for place recognition and goal localization. *Psychonomic Bulletin & Review, 14*(4), 676-680.

Valiquette, C. M., McNamara, T. P., & Labrecque, J. S. (2007). Biased representations of the spatial structure of navigable environments. *Psychological Research/Psychologische Forschung, 71*(3), 288-297.

Valiquette, C. M., McNamara, T. P., & Smith, K. (2003). Locomotion, incidental learning, and the selection of spatial reference systems. *Memory and Cognition, 31*(3), 479-489.

Waller, D. (2006). Egocentric and nonegocentric coding in memory for spatial layout: Evidence from scene recognition. *Memory & Cognition, 34*(3), 491-504.

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 867-882.

Wang, F. R., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition, 77*(3), 215-250.

Wang, F. R., & Spelke, E. S. (2002). Human spatial representation: insights form animals. *TRENDS in Cognitive Sciences, 6*(9), 376-382.