This is a pre-publication version of an article that appeared in the Journal of Experimental Psychology: Learning, Memory and Cognition, 2014, Vol 40, No. 2, 511-531.

One spatial map or many?

Spatial coding of connected environments.

Authors:

Xue Han [a] and Suzanna Becker [b]

a. Department of Psychology, Faculty of Education, Northeast Normal University

5268 Renmin Street, Changchun, Jilin, China, 130024

Phone: (86)431-85098069; Email: hanx100@nenu.edu.cn

Author to whom correspondence should be addressed

b. Department of Psychology, Neuroscience and Behaviour, McMaster University

1280 Main Street West, Hamilton, Ontario, Canada, L8S 4K1

Phone: 905-525-9140 x 23020, Fax: 905-529-6225; Email: becker@mcmaster.ca

Abstract

We investigated how humans encode large-scale spatial environments using a virtual taxi game. We hypothesized that if two connected neighborhoods are explored jointly, people will form a single integrated spatial representation of the town. On the other hand, if the neighborhoods are first learned separately and later observed to be connected, people will form separate spatial representations; this should incur an accuracy cost when inferring directions from one neighborhood to the other. Interestingly, our data instead suggest that people have a very strong tendency to form local representations, regardless of whether the neighborhoods were learned together or separately. Only when all visible distinctions between neighborhoods were removed did people behave as if they formed one integrated spatial representation. These data are broadly consistent with evidence from rodent hippocampal place cell recordings in connected boxes, and with hierarchical models of spatial coding.

Keywords: Spatial memory; Spatial map; Multi-scale hierarchical representations; Independent, local fragments; Multiple connected environments; Virtual reality.

Introduction

People solve spatial memory tasks using a variety of different strategies, likely employing multiple different internal representations. One such representation is the cognitive map (Tolman, 1948). While it is still debated whether non-human animals form some sort of integrated, map-like representation of space, it is widely accepted that humans have this ability. And yet, after many decades of research on this topic, the precise nature of our internal spatial representations remains a topic of considerable debate. One of many unresolved issues is when we construct a single spatial representation versus multiple local representations of large-scale spaces.

Some theories of spatial cognition have emphasized non-hierarchical, local representations. For example, Worden (1992) proposed that mammals store memories of their geographical environment as a collection of independent fragments, each consisting of a set of landmarks, their geometric relationships and their non-geometric properties. On the other hand, other theories have emphasized global representations. Within global theories, there are many possibilities, such as flat versus hierarchical representations. Some researchers postulate that people's spatial memories are organized hierarchically based on global and local properties of an environment (e.g. Hirtle & Jonides, 1985; McNamara, 1986; Meilinger, 2008; Stevens & Coupe, 1978). Similarly, Meilinger (2008) proposed a network of reference frames theory, which assumes that the space is encoded in multiple interconnected reference frames through the navigator's locomotion.

Local and global theories of spatial representation make different predictions about how people will combine information across spatial regions. If our representations are based on independent fragments, then integrating knowledge across fragments must occur at retrieval time, incurring a large cost in both accuracy and response time. In contrast, with a single flat representation, when judging a spatial relation between two points, people should be equally fast and accurate within versus between regions, assuming equal distance judgments. Finally, when using hierarchical representations, people should be able to employ their coarse-scale knowledge to infer directions from one region to another. In a hierarchical scheme, assuming that information is encoded more coarsely at larger spatial scales, when inferring directions from one region to another they must rely on coarser spatial information, thus there should be an accuracy cost. Whether there is also a speed cost depends upon the type of information encoded at each level. If objects are represented at a fine spatial scale but not at a coarser scale, inferring spatial relations between such objects should be slower when the objects are in different regions, requiring post-retrieval processes. On the other hand, for objects that are represented at all spatial scales, there should be no such speed cost, as participants can respond based on a coarse scale spanning both regions. Factors such as the object size, type of attention paid to objects, and their relevance to navigation may influence whether they are encoded locally or globally. Previous research has shown that when objects are relevant to navigation, they are more likely to be treated as landmarks (Han, Byrne, Kahana, & Becker, 2012) and encoded in the dorsal visuo-spatial pathway (Janzen & Turrenout, 2004). In the studies reported here, we probe people's memory for highly salient landmarks that have been visited frequently as navigational targets in a VR task. We therefore hypothesize that these landmarks will be represented at all spatial scales; if this is correct, then participants should be less accurate but equally fast at inferring directions between landmarks that span two regions. This would be consistent with the employment of multi-scale hierarchical representations in which the landmarks are represented at all levels of the hierarchy, albeit more coarsely at larger scales. On the other hand, if participants are both slower and less accurate when landmarks are in different regions, they may be using post-retrieval processes to integrate their spatial knowledge across regions. This would be consistent with either the employment of independent fragments, or a hierarchical representation in which the landmarks are represented at finer (within-region) but not coarser (between-region) spatial scales. These predictions are summarized in Table 1.

There is extensive empirical support for local representations, but much less evidence that would differentiate between independent fragments and hierarchical representations. For example, there is evidence that participants are more readily primed by locations in the same region than locations in different regions (Brockmole & Wang, 2002; McNamara, 1986), direction judgments between objects are affected by their superordinate spatial relations (McNamara, 1986), between-array geometry influences performance on within-array judgments (Greenauer & Waller, 2010), people make more errors in judging relations between landmarks located at widely separated geographical locations (Stevens & Coupe, 1978) and distances between objects in the same region tend to be underestimated , while distances between objects in different regions tend to be overestimated (McNamara, 1986). However, in many of these studies, the distances between objects within the same region versus across regions were not held constant (Greenauer and Waller, 2010; Stevens & Coupe, 1978), and reaction times for distance and direction judgments were not reported (Greenauer and Waller, 2010; McNamara, 1986; Stevens & Coupe, 1978). In the few studies that have reported reaction times for within versus between region spatial decisions while controlling for distance, evidence for hierarchical versus non-hierarchical theories is mixed. For example, participants were reported to have slower responses and larger errors for across-region pointing (Montello & Pick, 1993), consistent with the employment of independent fragments of local knowledge. On the other hand, in a spatially primed object recognition task, RTs were faster for same-region primes compared to between-region primes, but similar for same-region distant object primes versus different region close object primes (see Table 2, McNamara, 1986).

One factor that may be important in determining whether an environment is represented as one global representation or multiple independent local representations is whether the local regions were learned together or separately.   Evidence suggests than when two routes are learned on different levels of a building, pointing to landmarks across multiple levels of the building incurs a cost in both pointing latency and accuracy (Montello and Pick, 1993). The reaction time difference suggests that people may have employed post-retrieval strategies to integrate spatial information across multiple locally-constructed representations. It could be that combining spatial knowledge across three dimensions (multiple floors of a building) poses an unnatural barrier to spatial integration; within a two-dimensional environment, people may more readily integrate their spatial knowledge from local regions into a global perspective. However, similar findings have been reported in two-dimensional environments. When people were asked to estimate the directions of landmarks seen along an S-shaped route and a U-shaped route on a horizontal plane, they were less accurate in estimating landmarks when they were required to integrate their knowledge across the two routes, regardless of the distance from the current perspective to the landmark (Ishikawa and Montello, 2006). These data are consistent with either the construction of multiple local representations (independent fragments), or a hierarchy of representations within which landmarks are only represented at the local level.

Strong evidence for multi-scale spatial representations comes from electrophysiological recordings of place cells. Place cells, first discovered by O’Keefe and Dostrovsky (1971) and subsequently reported in many species including humans (Ekstrom et al., 2003), fire when the animal is within a specific local area and are often insensitive to heading direction (O'Keefe, 1976). Empirical evidence suggests that a place cell's spatial tuning is based on inputs from a multitude of cues, including the location of local boundaries from subicular “boundary vector cells”(Burgess, Jackson, Hartley, & O’Keefe, 2000; Hartley, Burgess, Lever, Cacucci, & O’Keefe, 2000; Lever, Burton, Jeewajee, O’Keefe, & Burgess, 2009), path integration cues from entorhinal grid cells (Hafting, Fyhn, Molden, Moser, & Moser, 2005), and contextual cues (Anderson & Jeffery, 2003). Moreover, recordings made in very large spaces suggest that place cells form the basis of a hierarchical or multi-scale representation, with some place fields spanning the entire length of an 18-meter track (Kjelstrup et al., 2008). Similarly, grid cells in the medial entorhinal cortex each fire at multiple locations arranged in a hexagonal grid (e.g. Fyhn, Molden, Witter, Moser, & Moser, 2004; Hafting et al., 2005; Sargolini et al., 2006), and their firing fields vary in spatial scale (Barry, Hayman, Burgess, & Jeffery, 2007).

Given the above evidence for local multi-scale representations of space, a key question is how larger regions of space are integrated, within very large-scale complex environments such as cities. One possibility is that local spatial representations are flexibly combined during the learning process into larger scale, more complex representations. If this is the case, then one would expect to see local representations of connecting paths between regions, and place cells that fire in one region or the other, but not both. Alternatively, separate representations might be formed for different local regions, and flexibly combined via post-retrieval processes. In the latter case, one would expect completely distinct sets of place cells firing in different sub-regions, and place cells with multiple unrelated firing fields in different regions. A wealth of electrophysiological evidence sheds light on this question. For example, in the hairpin maze, a complex environment with multiple turning points, both place cells and grid cells showed similar remapping patterns at the turning points (Derdikman et al., 2009), which suggests that regions separated by a turning point may be encoded as separate distinct representations; moreover, turning points may be encoded separately, providing a representational bridge between the different regions. Other evidence seems to support piecemeal, fragmented representations of space (Derdikman et al., 2009). For example, when two regions of an environment are learned separately, the place cells in the two environments bear no relation to each other and many cells have unrelated place fields in both regions (Tanila, 1999). Furthermore, opening a connection between the two environments causes many of the place cells to re-map and/or develop a single place field in just one part of the environment (Paz_Villagrán, Save, & Poucet, 2004), suggesting that the animal is treating the unified space as a new environment and generating a distinct internal representation in the latter case.

Based on the above evidence from human behavioural and animal electrophysiological studies, we predict that when there is local spatial structure in a large-scale environment, people will tend to construct multiple local representations, particularly when the multiple environments are explored separately. Whether the locally represented information is also represented at a coarser spatial scale may depend on a number of factors, including how the regions were explored and how salient the information is for orienting and navigation. Therefore, in our study, we had participants explore two connected virtual neighborhoods and investigated various conditions in which they might be encoded as one unitary spatial representation versus multiple piece-meal representations. In the latter case, combining two local representations together should incur a cost in accuracy when inferring directions from a location in one neighborhood to a location in the other neighborhood. Moreover, because we tested spatial memory for landmarks that were highly relevant to navigation, we predicted that they would be represented at multiple spatial scales. Therefore, if hierarchical representations are used, we predicted that people should be equally fast at making judgments between regions versus within regions. On the other hand, if independent fragments are employed, there should be a cost in reaction time for integrating across local representations at retrieval time.

Participants in the four experiments reported here learned the layout of two connected neighborhoods in a town by playing a virtual taxi game, after which they completed a memory task. The memory task required participants to point in the remembered direction of each of a set of target locations from viewpoints in each of the two neighborhoods. The target could either be in the same neighborhood as the test viewpoint or in the other neighborhood, with within- and between-neighborhood trials intermixed. Importantly, in these two types of trials, we controlled for the angles and distances between the viewer and the target locations. The targets for the memory tests were passenger drop-off (PDO) locations, which were expected to be well-learned, salient landmarks, given the goals of the virtual taxi game. In Experiment 1, participants initially learned the two neighborhoods separately; they were then shown a video clip of the pathway connecting the two neighborhoods, after which they explored the neighborhoods jointly. Because the results were most consistent with participants forming independent local representations (larger pointing latency and pointing errors for across neighborhoods pointing), in Experiment 2 participants were given more time to learn the town, and we varied the means by which they learned how the two neighborhoods were connected by allowing them to i) view, but not navigate, the connection between the neighborhoods; ii) view a video clip of being teleported along the connection; or iii) freely navigate along the connection. As in Experiment 1, within-neighborhood pointing to PDO locations was always more accurate than between-neighborhood pointing. However, in contrast to the results of Experiment 1, pointing latency was no different for within- versus between-region locations, suggesting that participants may have used hierarchical representations in Experiment 2. Experiment 3 examined the possibility that participants' between-neighborhood errors were due to an inability to accurately judge the length of the connection between neighborhoods by removing the fences that separated the two neighborhoods. Again participants were more accurate but no faster for within-neighborhood PDO locations. Finally, in Experiment 4 we removed all distinct features that differentiated the town's two neighborhoods. In this final experiment, there was no difference in memory between the two types of PDO locations in terms of either response speed or accuracy. Thus, participants were able to encode the large town as one single environment when there were no differentiating cues to spatially group the spatial features into local regions; in all other cases, their behaviour was more consistent with the construction of multiple local representations.

Experiment 1

After viewing images of the target and lure landmarks in a pre-exposure phase, participants implicitly learned the town layout and PDO locations by playing a virtual taxi game requiring active navigation. In subsequent spatial memory tests, they were asked to point in the remembered directions of well learned landmarks (the PDO locations) from cued viewpoints. Learning and spatial memory test phases were repeatedly interleaved in 5 blocks, taking a total of approximately 1 hour.

Method

Participants

Twenty-six McMaster University students (7 males and 19 females) of ages ranging from 18 to 32 years (mean age 19.31) participated in the experiment. Participants had normal or corrected-to-normal vision and received partial course credit for taking part in this experiment.

Written informed consent was obtained from all participants. This study was reviewed and approved by the McMaster Research Ethics Board.

Materials

We employed Kahana's "Yellow Cab" virtual driving simulator (see http://memory.psych.upenn.edu/Research) to simulate the virtual taxi game for the study phase of the experiment. Participants explored a single rectangular shaped town (21 by 10 VR units in size) consisting of two neighborhoods (see Figure 1a) connected by a navigable pathway that was initially occluded by an opaque, non-navigable barrier. Each of the two neighborhoods included eight distinctly textured and signed passenger drop off (PDO) locations as well as uniformly textured grey background buildings. One of the neighborhoods, hereafter referred to as neighborhood A (left side of the town, colored in purple in Figure 1a), was designed as a restaurant district, and the other, which we shall refer to as neighborhood B (right side of the town, colored in blue in Figure 1a), was designed as a shopping district. Each neighborhood was surrounded by distinctly colored and textured fences. To help participants encode the pathway connecting the two neighborhoods, in each neighborhood there was a distinct object located at the end of the connecting pathway that was visible from the other neighborhood. Each neighborhood had 9 distinctively marked buildings, including the 8 PDO locations and a building marking the starting points for passenger pickup trials that was never a PDO target. On memory test trials, the cued viewpoint was always an image of one of the two starting point buildings, "Mike's Restaurant" (see Figure 2a) in neighborhood A or "Aaron Chang Gallery" (see Figure 2b) in neighborhood B. 30 images of restaurants and shops were viewed in the pre-exposure phase, 18 of which appeared in the town and the remaining 12 of which served as distractors for the subsequent memory task. Within each neighborhood, the restaurants/shops were located in an irregular configuration with the following constraints (see Figure 1a): for a subset of six of the eight restaurants (numbered in Figure 1a) and the starting point in neighborhood A, the configuration of these 7 locations was identical to a mirror image of the locations of the corresponding 6 shops and starting point in neighborhood B. The remaining two restaurants and shops in each neighborhood were located irregularly to make the configurations asymmetric; however, only the responses to the symmetrically paired PDO locations were included in the analyses, in order to control for pointing distances and angles within versus between neighborhoods. The identities and locations of the restaurants and shops in the town remained constant across blocks. Participants were instructed to press and hold down the arrow keys on the keyboard to control their navigation, allowing them to turn in any direction, control their speed, or do a U-turn.

The memory test was similar to that used by Han et al (2012). It was implemented in Matlab with the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). On each memory test trial, a “navigator” appeared on the screen, consisting of a half compass shaped figure with an image of the town from the test viewpoint (one of the two starting points in the navigation phase) at the bottom of the navigator (see Figure 3a and 3.3b), and a rotatable red compass pointer. At the tip of the compass pointer there was an image of the target restaurant or shop for the current trial, which moved with the pointer. The target was always one of the 30 restaurants/shops shown in the pre-exposure phase, 18 of which had been in the town and 12 of which were distractors. The participant moved the mouse to rotate the compass pointer/target to the remembered direction of the target, and clicked the mouse to indicate their final response, or pressed the space bar to indicate that a landmark was not recognized as having been in the town.

Procedure

There was a total of five blocks. Blocks 1, 2, 4, and 5 each included a pre-exposure phase, a study phase and a test phase. Block 3 included of a pre-exposure phase, a video clip viewing phase and a test phase. In the pre-exposure phase, each of 30 restaurants and shop images appeared for two seconds followed by a blank screen for one second. The purpose of this pre-exposure phase was to establish a degree of familiarity with the distractors for the subsequent recognition memory test. Participants were informed that some of the restaurants and shops would appear in the town. In each study phase, the participant was instructed to complete five passenger pickup trials. For each of these trials, the participant was asked to act as a taxi driver, roam through the neighborhood to find a randomly located passenger, and deliver the passenger to the requested location. Passenger drop-off locations (PDOs) were always restaurants or shops, which were visibly distinct, clearly signed locations in the town. A passenger pickup trial began with the participant located at one of two starting points, facing toward the town. He/she was asked to navigate freely until a passenger was found and "collected" by bumping into the passenger. A textual cue to the goal location then appeared, e.g. "Please take me to the Computer Store, I will give you 100 points", and the participant was required to navigate as quickly and efficiently as possible to drop off the passenger to the goal location by bumping into it.

In Block 1, participants explored only half of the town, while in Block 2 participants explored the other half. Which of the two neighborhoods was explored first was counterbalanced between participants. In the first two blocks, there was no visible or accessible connection between the two neighborhoods. In Block 3, there was no active navigation; instead, after completing the pre-exposure phase, participants were told that they would view a video clip illustrating how the two neighborhoods were connected. The video showed a trajectory of driving from one starting point to the other (from starting point A to B for half the participants and in the reverse direction for the other half). In Blocks 4 and 5, the participants were told that the visible barrier between the two neighborhoods would be removed and the connection would be open and navigable, allowing them to travel freely back and forth between the two neighborhoods and explore the entire space of the town while locating and delivering passengers.

In each of the active navigation blocks (1, 2, 4 and 5), the study phase was terminated when either the participant had successfully found and delivered five passengers or 10 minutes had elapsed. In Blocks 1 and 2, after each passenger delivery, the participant was relocated to the same starting point within the neighborhood to begin the next passenger pickup. In the final blocks 4 and 5 (after removal of the barrier) he or she was relocated to the starting point in the adjacent neighborhood alternatingly between pickups.

Immediately after each study phase, participants completed a simultaneous test of recognition memory and spatial memory. The test required participants to point in the direction of the remembered location of each PDO building from a given viewpoint as quickly as possible, with no feedback. In Blocks 1 and 2, the tested viewpoint was always an image of the starting point in the same neighborhood explored in the preceding study block. In Blocks 3, 4 and 5, spatial memory for each PDO was tested twice, once from each of the two tested viewpoints/starting points (see Figure 3a and b). On each test trial, if the participant thought the restaurant or shop had not appeared in the town, he or she pressed the ‘space bar’, and the next restaurant or shop would be displayed. Otherwise, he or she pointed in the remembered direction of the PDO location from the displayed viewpoint by using the mouse to move the compass pointer in the desired direction, and then pressing the left mouse button[1]. After each response, the

[1] Note: We did not measure recognition memory reaction time separately, but we did

next target restaurant or shop would be displayed on the compass pointer. The order of presentation of the targets and tested viewpoints was randomized within each block for each participant. The recognition responses, pointing directions and total reaction time for the combined spatial/recognition memory response were recorded during the memory test phase. After 5 blocks of study and test phases, participants were asked to draw a map of the town, including the two neighborhoods, restaurants and shops, on a blank piece of paper.

Data Analysis

Participants' data were excluded from further analyses if they misaligned the two neighborhoods on the final map-drawing task, indicating that they failed to learn the overall town layout.

Within-neighborhood pointing responses were defined as responses made from the viewpoint of starting point A to restaurants 1a, 2a, and 3a in neighborhood A, and from starting point B to shops 4b, 5b, and 6b in neighborhood B (see Figure 1). Between-neighborhood pointing responses were defined as responses from the viewpoint of starting point A to shops 1b, 2b, and 3b in neighborhood B, and from starting point B to restaurants 4a, 5a, and 6a in neighborhood A.

In a preliminary analysis, to verify that spatial memory tests for the shop and restaurant neighborhoods were matched for difficulty, the recognition accuracy, pointing latency and pointing errors (which were recorded prior to participants being shown the connection between the two neighborhoods) for the 6 matched restaurants and shops in each neighborhood were compared across the two neighborhoods.

---

measure pointing latency, as our memory test combined recognition and spatial memory.

Preliminary analyses included Block as a factor for all four experiments. Not surprisingly, significant improvement across blocks was seen in terms of both pointing responses and recognition memory. More specifically, pointing latency and recognition memory accuracy improved significantly over blocks in all four experiments, and pointing errors improved significantly in all cases except Experiment 1. However, to simplify the presentation of results, block effects are not reported except in cases where a Block by Location (pointing to within- versus between-neighborhood landmarks) interaction was significant. Bonferroni corrections were used for all multiple comparisons throughout this paper.

Recognition Accuracy

A recognition response was counted as correct if the participant made a pointing response in any direction to a restaurant or shop that had appeared in the town, or pressed the space bar for any lure image that had not appeared in the town. We calculated percent correct recognition separately for pointing to PDO locations within- and between-neighborhoods. A two-tailed paired sample t-test was used to compare the difference in recognition memory accuracies between the two types of PDO locations.

Pointing Latency

There were two reaction times (in seconds) for each PDO location in each block: one for each of the two tested viewpoints. Reaction times for correct responses, averaged across blocks and tested viewpoints, were compared for pointing to within- and between-neighborhood PDO locations using a two-tailed paired sample t-test.

Pointing Error (Average Absolute Pointing Errors)

Each participant had two pointing responses for each PDO location, one from the viewpoint of “Mike’s Restaurant” and one from the viewpoint of “Aaron Chang Gallery”, in each block. A pointing error was calculated as the absolute value, in degrees, of the difference between the pointing response direction and the PDO location's actual direction. For each participant, average pointing errors were calculated separately for within- and between-neighborhood pointing trials. For each of the two trial types, for each cued viewpoint there were six pointing errors in each block (assuming that participants correctly recognized all PDO locations); these were averaged across blocks and across viewpoints. A two-tailed paired sample t-test was used to compare the difference in pointing errors between the two types of PDO locations (within versus between neighborhoods).

Results

Based on their performance on the map drawing task, six participants (5 females and 1 male) failed to learn the spatial relationship between the neighborhoods; their data were excluded on that basis, leaving data from twenty participants (6 males and 14 females) for further analysis.

To check for differences in encoding difficulty between the two neighborhoods, we performed a preliminary analysis of the data from Blocks 1 and 2. This analysis revealed that there was no difference in spatial memory for buildings in the two neighborhoods in terms of recognition accuracy (t=1.633, df=19, p=0.119, Cohen's d=0.51; Restaurant neighborhood: mean=92.50%, SE=2.6%; Shopping neighborhood: mean=85.83%, SE=3.3%), pointing latency (t=0.554, df=19, p=0.586, Cohen's d=0.13; Restaurant neighborhood: mean=5.45, SE=0.49; Shopping neighborhood: mean=5.17, SE=0.45) or pointing errors (t=0.065, df=19, p=0.949, Cohen's d=0.02; Restaurant neighborhood: mean=34.02, SE=2.79; Shopping neighborhood: mean=33.75, SE=3.07). In subsequent analyses, we averaged data across all study blocks.

Recognition accuracy

A two-tailed paired sample t-test of the recognition accuracy revealed no significant difference between the two types of PDO locations (*t*=1.65, *df*=19, *p*=0.116, Cohen's *d*=0.42; Within neighborhood locations: mean=95.56%, SE=1.2%; Between neighborhoods locations: mean=92.78%, SE=1.7%), see Table 5.

Pointing latency

A two-way repeated measures Location x Block ANOVA of the pointing latencies revealed significant main effects of Location [F(1, 19)=43.643, *p*<0.001, Observed Power=1] and Block [F(2, 38)=19.772, *p*<0.001, Observed Power=0.996], and a significant interaction between Location and Block [F(2,38)=3.825, *p*=0.032, Observed Power=0.653]. To investigate the Block by Location interaction, separate two-tailed paired-sample *t*-tests were used to compare within- and between-neighborhood responses at each block; these revealed that pointing latency was significantly faster for within neighborhood than for between neighborhood PDO locations at each block (Block 3: *t*=-4.76, *df*=19, *p*<0.001; Block 4: *t*=-3.499, *df*=19, *p*=0.002; Block 5: *t*=-5.259, *df*=19, *p*<0.001). Because the location effect was significant at every block, pointing latencies were averaged across blocks for further analysis.

Averaged across blocks, a two-tailed paired sample t-test of the pointing latencies for within- versus between-neighborhood landmarks revealed a significant difference between the two types of PDO locations (*t*=-6.61, *df*=19, *p*<0.001, Cohen's *d*=-0.76). Responses were faster for pointing to locations within neighborhoods (mean=2.87, SE=0.24) than between neighborhoods (mean=3.76, SE=0.29), see Figure 4 and Table 5.

Pointing errors

A two-tailed paired sample t-test of the pointing errors averaged across blocks revealed a significant difference between the two types of PDO locations (*t*=-4.18, *df*=19, *p*=0.001, Cohen's *d*=-1.06). Errors were significantly smaller when pointing to PDO locations within neighborhoods (mean=24.49, SE=1.72) than between neighborhoods (mean=37.74, SE=3.57), see Figure 5 and Table 5.

Discussion

When the two neighborhoods were first learned separately, and later explored jointly, participants behaved as if they constructed separate spatial representations of the two neighborhoods. We hypothesized that combining two local representations would result in reduced accuracy and increased latency of between-neighborhood pointing responses if independent representations were formed, both of which were seen in our results. Pointing responses for PDO locations in the same neighborhoods as the testing viewpoints were faster than for those in the adjacent neighborhoods, even when we only analyzed PDO locations that were, on average, equi-distant from the observer (i.e. each restaurant or shop within the same neighborhood was paired with a shop or restaurant at an equal distance and angle away from the observer in the other neighborhood). The longer time they took did not improve their accuracies for PDO locations in the adjacent neighborhoods; on the contrary, pointing errors were larger for those PDO locations. Interestingly, while the latency differences appeared to be dissipating across blocks (Figure 4), the accuracy differences showed a trend toward increasing across blocks (Figure 5). This could mean that as learning proceeds, participants undergo a transition from separate, independent representations to integrated hierarchical representations. Another possibility is that participants formed a hierarchical representation of the town throughout the experiment, but the PDO locations were only represented at the level of local neighborhoods, and not at a coarser spatial scale spanning the entire town. When required to point to a PDO location in a different neighborhood, their coarse level would not be of any use for this task, forcing them to integrate their local representations of buildings in each neighborhood at retrieval time. The latter interpretation would not readily explain why the reaction time differences decreased across blocks.

The two neighborhoods may have been treated as separate environments because they were explored separately at the beginning of the experiment. If instead the two neighborhoods had been explored together from the beginning, they might have been treated as a unified environment. Moreover, as noted above, the difference in pointing latency between the two types of locations was shrinking across blocks (Figure 4). Moreover, in different phases of Experiment 1, participants actually experienced the two connected neighborhoods in three different ways: first separately, then by passively viewing the connection between the neighborhoods, and finally, via joint exploration. It is possible that with more learning time, and the opportunity to explore the two neighborhoods together from the outset, participants may have been able to integrate their knowledge of the two neighborhoods into a global representation. Another factor that may be important in determining whether two regions are integrated in spatial memory is the manner in which they are explored, whether by direct experience or by simply observing the way regions are connected.

To tease apart the effect of learning by passive viewing versus active navigation on the formation of integrated spatial representations, in Experiment 2 we allowed them a greater number of blocks to explore the two neighborhoods, and varied the means by which they learned how the two neighborhoods were connected. In a view-only condition, they could look along the pathway connecting one neighborhood to the other but could not navigate along it. In a teleport condition, they passively viewed a movie of the trajectory along the connecting pathway. Finally, in a “whole town” condition, they could navigate freely along the pathway in either direction.

Experiment 2

In this experiment, each participant was pseudo-randomly assigned to one of three conditions: View, Teleport, and Whole. This was a 2-hour experiment with learning and spatial memory test phases repeatedly interleaved in 6 blocks.

Method

Participants

Sixty McMaster University students participated in the experiment; age ranged from 18 to 37 years, and the mean was 20.02. There were 20 participants in each condition (10 males and 10 females). Participants had normal or corrected-to-normal vision. Participants received two course credits or $20 for taking part in this experiment. This study was reviewed and approved by the McMaster Research Ethics Board. Written informed consent was obtained from all participants.

Materials

The materials for the VR component of the study were the same as those used in Experiment 1. In addition, we had participants complete a questionnaire adapted from the Santa Barbara Sense of Direction Scale (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002; see Appendix A) to investigate what factors may be associated with performance on the pointing task. We added two questions to the original SBSOD questionnaire. Question 16 had participants rate on a 7-point scale how likely they were to rely on a GPS when they travel to new places, while Question 17 asked participants whether they were video game players (a yes/no question).

Procedure

The experiment had six blocks, each consisting of a pre-exposure phase as in Experiment 1, a study phase and a test phase. Each study phase used similar procedures to those of Experiment 1, except that there were three conditions, which varied between subjects. In the *View* condition, in the first half of each study phase, participants explored one of the two neighborhoods but could see the other through a visible but non-navigable pathway connecting the two neighborhoods. In the second half of each study phase, participants explored the other neighborhood, and again could see the first neighborhood along the connecting path but could not cross through it. The starting neighborhood was counterbalanced between participants. Throughout the study phase, after each passenger delivery, participants were relocated to the starting point in the same neighborhood to start another passenger pick-up, until they had successfully found and delivered five passengers within that neighborhood, after which they completed five passenger pickup trials in the other neighborhood.

The *Teleport* condition was similar to the View condition except that the pathway along the connection between neighborhoods was neither visible nor accessible. In each block, after learning both neighborhoods separately, the participants watched a video clip showing how the two neighborhoods were connected: moving from starting point A to starting point B or from B to A. The video clips and starting neighborhoods were counterbalanced between participants.

In the *Whole* condition, the pathway connection was visible and navigable throughout the experiment, and the starting neighborhood alternated between passenger pickup trials. Thus, each time a passenger was delivered, the participant was re-located to the starting point in the other neighborhood, facing either “Mike’s Restaurant” or “Aaron Chang Gallery”, before being cued to collect the next passenger. Unlike in the View and Teleport conditions, the passenger and/or PDO location might be either in the same neighborhood as the starting point or in the other neighborhood, thus requiring the participant to explore both neighborhoods jointly.

In all three conditions, which neighborhood was explored first was counterbalanced between participants and there were ten passenger deliveries in each of the six blocks, hence a total of 60 passenger deliveries.

The same memory test and mapping task were used as in the last three blocks in Experiment 1. Additionally, at the end of the final block, participants were asked to answer the Santa Barbara sense of direction (SBSOD) questionnaire, augmented with questions concerning their GPS usage and video game playing (see Appendix).

Data Analysis

Analyses were the same as those used as in Experiment 1, except we had three conditions. Therefore, we conducted two-way repeated measures ANOVAs, with Location (within vs. between neighborhoods) as a within-subject factor and Condition (View vs. Teleport vs. Whole) as a between-subject factor. We also separately analyzed the final performance in the last block to investigate whether, with more learning time, the two neighborhoods would eventually be treated as one environment.

In addition, based on the questionnaire results, we investigated individual differences in spatial abilities via correlational analyses. Answers to Questions 1 to 15 (the questions in the original version of the SBSOD) were coded such that larger values indicated a better sense of direction. These sum of ratings yielded final SBSOD scores. Correlations were calculated between SBSOD scores and the three memory measures: recognition accuracy, pointing latency and pointing errors. Questions 16 and 17 were analyzed separately. Question 16 had participants rate on a 7-point scale how likely they were to rely on a GPS when they travel to new places, while Question 17 asked participants whether they were video game players (a yes/no question).

We also compared SBSOD scores of participants who correctly mapped the two starting points with those who did not, to assess the validity of the map-drawing task.

Results

Based on the map-drawing task, four participants (2 females and 2 males) in the View condition, three (2 females and 1 male) in the Teleport condition, and eight (6 females and 2 males) in the Whole condition incorrectly aligned the two neighborhoods; their data were excluded from further analyses of spatial memory measures.

Recognition accuracy

A two-way repeated measures Location x Condition ANOVA of recognition memory accuracy revealed no significant effects or interactions (Location [$F(1, 42)=3.85$, $p=0.057$, Partial Eta Squared=0.084]; Condition [$F(2, 42)=0.48$, $p=0.624$, Partial Eta Squared=0.022], Location and Condition [$F(2, 42)=3.09$, $p=0.056$, Partial Eta Squared=0.128], (Within-Neighborhood PDO locations: View: mean=96.64%, STD=3.9%; Teleport: mean=97.44%, STD=4.1%; Whole: mean=98.27%, STD=2.2%; Between-Neighborhood PDO locations: View: mean=97.71%, STD=2.5%; Teleport: mean=98.73%, STD=2.2%; Whole: mean=97.74%, STD=2.5%).

Pointing latency

A two-way repeated measures Location x Condition ANOVA of the pointing latencies revealed a main effect of Condition [$F(2, 42)=3.87$, $p=0.029$, Partial Eta Squared=0.156], but no effect of Location [$F(1, 42)=0.062$, $p=0.805$, Partial Eta Squared=0.001] nor any interaction between Condition and Location [$F(2, 42)=0.85$, $p=0.434$, Partial Eta Squared=0.039], see Figure 6. Post hoc Bonferroni-corrected t-tests showed that pointing responses were significantly faster in the Whole condition than in View condition ($p=0.033$), but no other pairwise differences were significant (View: mean=4.23, SE=0.34; Teleport: mean=4.45, SE=0.33; Whole: mean=3.07, SE=0.40).

Pointing errors

A two-way repeated measures Location x Condition ANOVA revealed main effects of Location [$F(1, 42)=29.95$, $p<0.001$, Partial Eta Squared=0.416] and Condition [$F(2, 42)=4.34$, $p=0.019$, Partial Eta Squared=0.171], but no interaction between Condition and Location [$F(2, 42)=0.58$, $p=0.565$, Partial Eta Squared=0.027]. Post hoc comparisons showed that between-neighborhood pointing errors (mean=41.48, SE=1.99) were significantly larger than within-neighborhood errors (mean=29.84, SE=1.97). Pointing errors in the Whole condition (mean=42.64, SE=3.20) were significantly larger than those in the View condition ($p=0.017$, mean=30.34, SE=2.77), while there was no difference in responses between the View and Teleport conditions ($p=1.00$) or between the Teleport and Whole conditions ($p=0.135$). For details of pointing errors by blocks, see Figure 7.

Final performance

In all three measures, participants' performance improved over blocks (e.g. see Figure 5 for pointing errors). Given the trend in Experiment 1 for the Location effect to decrease across blocks, we therefore investigated whether the Location effect dissipated after 6 blocks of learning had taken place, by analyzing all three memory measures in the final block using two-way repeated measures Location x Condition ANOVAs. As in the main analysis, there were no significant main effects or interactions in terms of recognition accuracy or pointing latency. However, in terms of pointing errors, even in the final block of learning, there was a main effect of Location [$F(1, 42)=26.07$, $p<0.001$, Partial Eta Squared=0.383], but no effect of Condition [$F(2, 42)=2.34$, $p=0.109$, Partial Eta Squared=0.10] and no interaction between Location and Condition [$F(2, 42)=0.128$, $p=0.88$, Partial Eta Squared=0.006] (see Figure 7). As in the overall analysis across all blocks, in the final block, pointing within neighborhoods was more accurate (mean error=22.65, SE=1.87) than pointing between neighborhoods (mean error=36.71, SE=2.36) across conditions.

Questionnaire results

T-tests revealed that participants who correctly mapped the two starting points had higher SBSOD scores (t=2.17, df=58, p=0.039, Cohen's d=0.67) and recalled more stores in the mapping task (t=3.372, df=58, p=0.001, Cohen's d=0.92) than those who did not. Moreover, correlational analyses of all participants' data (including those who did and did not correctly map the starting points) revealed that SBSOD scores were positively correlated with the number of stores recalled in the mapping task (r(60)=0.307, p=0.017). This significant correlation with a widely used test of spatial abilities provides validation of our use of the pointing task and map-drawing task as measures of spatial ability.

Video Game Players

Across conditions, participants with video game experience (gamers) (n=27, mean=16.04, SE=0.47) recalled significantly more restaurants and shops in the mapping task than those without video game experience (non-gamers) (n=18, mean=13.83, SE=0.8) (t=2.53, df=43, p=0.015, Cohen's d=0.74), but were no better in terms of recognition accuracy, pointing latency, or pointing errors. Thus, video game experience translated into better recognition memory for landmarks but not into better judgments of the directions of landmarks within or between neighborhoods.

View condition

In the View condition, SBSOD scores were not significantly correlated with any of the measurements (see Table 2). GPS score (Q16, on a 7-point scale, where a high score indicates the participant is LESS likely to use a GPS when travelling to new places) positively correlated with recognition accuracy (*r*(16)=0.533, *p*=0.033) and negatively correlated with pointing errors (*r*(16)=-0.555, *p*=0.026) when pointing between neighborhoods. Thus people who relied less on a GPS in daily life were more accurate at integrating their knowledge between neighborhoods.

Teleport condition

In the Teleport condition, SBSOD scores were significantly negatively related to pointing errors for within-neighborhood PDO locations, *r*(16)= -0.718, *p*=0.001 (see Table 2). In other words, people with a better sense of direction performed more accurately on the within-neighborhood trials. However, there were no significant correlations between any of the other measures and the SBSOD scores, nor was GPS usage correlated with any of the measurements.

Whole condition

In the Whole condition, the SBSOD scores were significantly negatively correlated with pointing latency for between-neighborhood PDO locations, *r*(12)=-0.591, *p*=0.043, and marginally negatively correlated pointing latency for within-neighborhood PDO locations, *r*(12)=-0.572, *p*=0.052 (see Table 2). No other correlations with between the SBSOD scores were significant in this condition (see Table 2), nor were there any significant correlations between GPS usage and any of the memory measures in the Whole condition. Thus, people with a better sense of direction, when forced to learn the two neighborhoods together, were faster but no more accurate at judging directions within and between neighborhoods.

Discussion

In all three conditions, regardless of whether the two neighborhoods were explored separately or jointly, participants were more accurate at pointing to PDO locations within neighborhoods than between neighborhoods. This difference persisted even in the final block, in all three conditions, suggesting that even after extensive learning, the two neighborhoods were still encoded as local representations. However, in contrast to our findings in Experiment 1, there were no pointing latency differences in any of the three conditions in Experiment 2. If people were constructing independent local representations, and were forced to integrate their knowledge at retrieval time, we predicted that there would be a reaction time cost. Thus, our results suggest instead that participants were using hierarchical representations, with fine scale local representations for each neighborhood and a coarser scale global representation for the whole town. One difference between these two experiments that might result in different strategies was the exploration time. In Experiment 1, participants had to complete a total of 20 passenger pick-ups, taking less than an hour to complete the experiment, as compared to 60 pick-ups in Experiment 2, which took two hours to complete. It would be of interest for future research to further investigate the effect of learning experience on the types of representations people form. It is possible that when local regions are initially explored, independent, fragmented representations are formed, whereas with greater experience people eventually form either hierarchical multi-scale or flat unitary representations.

There were clear individual differences in how participants carried out the pointing task. For example, in the View condition, participants who relied less on a GPS in daily life were better at pointing to PDO locations between neighborhoods. The View condition poses the greatest challenge to participants in integrating their knowledge of the two neighborhoods. In contrast to the Teleport and Whole conditions, participants never visually experience moving along the corridor to see how the neighborhoods are connected. They have to piece the two neighborhoods together without actually either actively or passively navigating between them. Those who do not rely on a GPS may be more adept at visuo-spatial imagery, and therefore better able to imagine moving along the connecting pathway without having directly experienced it.

Interestingly, whereas GPS use negatively predicted performance in the View condition, sense of direction scores positively predicted performance in the Teleport and Whole conditions. Those who had higher SBSOD scores had the advantage in both conditions, as evidenced by greater pointing accuracy in the Teleport condition and shorter pointing latencies in the Whole condition. It is possible that these two conditions favor two distinctly different response strategies, a topic for future research.

Although the results of this experiment are consistent with the notion that participants treated the two neighborhoods as two separate environments in all three conditions, an alternative explanation is that participants formed flat global representations of the two neighborhoods (or could precisely combine the two separate representations), but made errors in judging the *length* of the pathway connecting the two neighborhoods, resulting in larger errors in pointing to PDO locations between neighborhoods. This alternate hypothesis is weakened by the observation that participants in all three conditions exhibited the same PDO location effect even though they had different experiences with the connection between the neighborhoods. Nevertheless, to rule out this alternative explanation, in Experiment 3, the fences between the two neighborhoods were removed while all other features of the neighborhoods remained the same. As a result, the possibility of participants misjudging the length of the pathway should be reduced, and therefore, pointing errors for within- and between-neighborhood PDO locations ought to be the same if they are using one flat representation.

Experiment 3

In Experiment 3, in order to encourage participants to form a single global representation of the environment, they explored both neighborhoods jointly from the very beginning, as in the

“Whole” condition in Experiment 2. The fences separating the two neighborhoods were removed (see Figure 1b) but the town still consisted of two visually and semantically distinct neighborhoods, a Restaurant district and a Shopping district, each surrounded by differently colored fences. Therefore, we hypothesized that errors in pointing to PDO locations within-neighborhood would be smaller than errors between-neighborhoods. This result would suggest that the pointing error difference observed in Experiment 2 was not due to misjudgments of the length of the pathway, but rather to participants basing their responses on two separate local representations.

Method

Participants

Twenty McMaster University students (9 males and 11 females) participated in this experiment; age ranged from 18 to 30 years, and the mean age was 20.9. Participants had normal or corrected-to-normal vision. Participants received either two course credits or $20 for taking part in this 2-hour experiment. This study was reviewed and approved by the McMaster Research Ethics Board. Written informed consent was obtained from all participants involved in this study.

Materials

The materials were the same as those used in Experiment 2, except that the fences along the pathway connecting the two neighborhoods were removed (see Figure 1b). Therefore, there was no distinct pathway connecting the two neighborhoods. However, the two neighborhoods were still visually and semantically distinct; the fences surrounded each neighborhood differed in color and texture, and one neighborhood contained only restaurants while the other contained only shops.

Procedure & Data Analysis

The procedures and analyses were the same as those used in the Whole condition in Experiment 2. All participants successfully drew the map.

Results

Recognition accuracy

A two-way repeated measures Block by Location ANOVA revealed significant main effects of Block [$F(5, 95)=10.086$, $p<0.001$, Observed Power=1] and Location [$F(1, 19)=6.603$, p=0..019, Observed Power=0.684] and a significant interaction between Location and Block [$F(5, 95)=4.26$, $p=0.002$, Observed Power=0.953]. Post hoc comparisons showed that recognition accuracy was significantly worse ($p=0.019$) for PDO locations within the same neighborhood (mean=96.7%, SE=0.8%) than for those in the adjacent neighborhood (mean=98.6%, SE=0.5%) and was significantly worse in Block 1 than in Blocks 3 ($p=0.01$), 4 ($p=0.031$), 5 ($p=0.024$), and 6 ($p=0.009$). Paired-sample t-tests at each block, corrected for multiple comparisons, revealed no significant differences between the two types of locations in any of the blocks.

Pointing latency

A two-tailed paired sample t-test of the pointing latency revealed no significant difference between pointing to PDO locations within versus between neighborhoods ($t=0.122$, $df=19$, $p=0.904$, Cohen's $d=0.02$; within-neighborhood PDO locations: mean=3.50, SE=0.21; between-neighborhood PDO locations: mean=3.48, SE=0.24), see Figure 8.

Pointing errors

A two-tailed paired sample t-test of the pointing errors revealed a significant difference between the two types of PDO locations ($t=-14.00$, $df=19$, $p<0.001$, Cohen's $d=-2.59$); errors were smaller for pointing to PDO locations within neighborhoods (mean=22.20, SE=1.78) than between neighborhoods (mean=41.95, SE=1.63). For details of pointing errors by blocks, see Figure 9.

Questionnaire results

Participants with video game experience (n=10) did not differ from non-gamers (n=10) in terms of recognition accuracy, pointing latency, pointing errors, or number of restaurants and shops recalled in the map-drawing task.

None of the spatial memory measures were significantly correlated with the SBSOD scores (see Table 3). GPS usage was significantly negatively correlated with pointing latency for between-neighborhood PDO locations, $r(20)= -0.502$, $p=0.024$. Thus, less reliance on a GPS in daily life was associated with faster pointing responses for between-neighborhood PDO locations.

Discussion

The results were similar to those obtained in the Whole condition in Experiment 2: pointing to PDO locations was more accurate within neighborhoods than between neighborhoods, even in the last block (see Figure 7), and even in the face of better recognition memory for between-neighborhood PDO locations. These findings argue against the alternative hypothesis that larger between-neighborhood pointing errors were due to participants misjudging the length of the pathway connecting the two neighborhoods. Even without fences to separate the two neighborhoods, participants still behaved as if they treated them as two separately and hierarchically encoded environments. Moreover, participants who relied on a GPS more often in their daily lives were slower at pointing to PDO locations in the adjacent neighborhoods.

One possible alternative explanation for the current findings is that the environment simply was too big to be encoded as one representation. Alternatively, the two types of PDO locations may have differed in important ways. For example, it is possible that the difference in pointing errors reflects the fact that all within-neighborhood PDO locations were located around the edge of the town whereas between-neighborhood PDO locations were located in the centre of the town. To rule out these possibilities, we conducted a final experiment in which the types of PDO locations were mixed between the two neighborhoods and the environment no longer contained distinct boundaries delineating the two neighborhoods.

Experiment 4

Experiment 4 was identical to Experiment 3 except that the distinctly coloured fences surrounding the two neighborhoods were replaced by uniformly textured walls and the restaurants and shops were intermixed within the town. Therefore, in contrast to the three previous experiments, there were no spatial, visual or semantic features to differentiate the two neighborhoods. Nevertheless, we still analyzed the same 6 pairs of PDO locations (see Figure 1c) as in the previous experiments. For cross-experiment comparison purposes, we therefore still refer to them as within- and between-neighborhood PDO locations. We hypothesized that there would be no difference in pointing responses for within and between neighborhood PDO locations. This result would suggest that the large pointing error differences observed in Experiments 1-3 were caused by participants constructing and using separate representations of the distinct neighborhoods, whenever there were features available to differentiate and cluster local landmarks into sub-regions.

Method

Participants

Twenty McMaster University students (11 males and 9 females) participated in the experiment; age ranged from 18 to 35 years, and the mean age was 20.6. Participants had normal or corrected-to-normal vision, and received either two course credits or $20 for taking part in this 2-hour experiment. This study was reviewed and approved by the McMaster Research Ethics Board. Written informed consent was obtained from all participants involved in this study.

Materials

The materials were the same as those used in Experiment 3, except for the following changes to the town: The fences distinguishing the two neighborhoods were removed, and restaurants and shops were intermixed across the town rather than being clustered by category within one or the other neighborhood (see Figure 1c). Therefore, there was no pathway connecting the two neighborhoods, and there were no other distinctions between the two neighborhoods. In this case, therefore, the town should be perceived as one environment.

Procedure & Data Analysis

The procedures and analyses were the same as those used in Experiment 3. All participants successfully drew the map.

Results

Recognition accuracy

A two-tailed paired sample t-test of the recognition accuracy revealed no significant difference between the two types of PDO locations (t=-1.19, df=19, p=0.249, Cohen's d=-0.25; within-neighborhood PDO locations: mean=96.53%, SE=0.8%; between-neighborhood PDO locations: mean=97.36%, SE=0.7%).

Pointing latency

A two-tailed paired sample t-test of the pointing latency revealed no significant difference between the two types of PDO locations (t=1.43, df=19, p=0.169, Cohen's d=0.16; within-neighborhood PDO locations: mean=3.91, SE=0.33; between-neighborhood PDO locations:

mean=3.69, SE=0.30), see Figure 10.

Pointing errors

A two-tailed paired sample t-test of the pointing errors revealed no significant difference between the two types of PDO locations (t=-1.74, df=19, p=0.098, Cohen's d=-0.33; within-neighborhood PDO locations: mean=35.25, SE=3.16; between-neighborhood PDO locations: mean=39.29, SE=2.21) (see Figure 11).

Questionnaire results

Participants with video game experience (n=12) did not differ from non-gamers (n=8) on any of the performance measures.

None of the memory measures correlated significantly with either SBSOD scores or GPS usage (see Table 3).

Discussion

Unlike the previous three experiments, in Experiment 4 there was no indication that errors depended on whether participants were pointing within- or between-neighborhoods, suggesting that participants may have encoded the large scale environment within a single flat spatial representation. Although the difference in pointing errors for within and between-neighborhood trials was not significant (p=.098), it was in same the direction as in the previous three experiments. This could be due to the positions of two types of locations in the town. In our study, the between-neighborhood PDO locations were arranged in the middle of the town and the within-neighborhood PDO locations were arranged closer to the edges of the town (see Figure 1). It could be that it is more difficult to estimate the directions of buildings that are closer in the middle of the town than those buildings that are closer to the edges. However, this does not explain why in the previous three experiments we observed very large, highly significant differences in within- versus between-neighborhood pointing errors relative to the very small, non-significant difference observed here.

Cross Experimental Comparisons

Our main question in the studies reported was whether there would be latency and/or accuracy differences when participants tried to remember the locations of landmarks in the same versus in a different neighborhood from the tested viewpoint. We found a location effect (a difference between within- and between-neighborhood pointing trials) on response latency in Experiment1 but in no other experiment. On the other hand, we found a location effect on response accuracy in Experiments 1-3 but not in Experiment 4. According to our predictions laid out in Table 1, these results are consistent with the use of independent representations of the two neighborhoods in Experiment 1 and integrated representations in Experiments 2-4, and the use of hierarchical representations in Experiments 2-3 but not in Experiment 4. Alternatively there may have been other differences between the experiments that meant the difficulty of within-neighborhood pointing trials varied, in the absence of qualitative differences in the type of representation participants may have used. To further explore this possibility, cross-experiment comparisons were made between Experiments 1 (blocks 3, 4 and 5), 2 (Whole condition), 3 and 4 on both pointing latency and accuracy.

Pointing Latency

To investigate the possibility that the spatial learning problem faced by participants was somehow easier in Experiment 1 relative to Experiments 2-4 , resulting in longer within-neighborhood response latencies in the latter experiments, a cross-experimental comparison was made using a two-way repeated measures Location x Experiment ANOVA of the pointing latencies. This analysis revealed a significant interaction between Location and Experiment [F(3, 68)=13.003, *p*<0.001, Partial Eta Squared=0.365], but no main effects of Location [F(1, 68)=1.489, *p*=0.227, Partial Eta Squared=0.021] or Experiment [F(3, 68)=1.093, *p*=0.358, Partial Eta Squared=0.046], see Figure 13. Separate post-hoc one-way ANOVAs of pointing latencies on within- and between-neighborhood trials revealed that response speed did not differ across experiments for between-neighborhood trials [F(3,71)=1.29, *p*=0.285, Partial Eta Squared=0.054], but was marginally significantly different across experiments for within-neighborhood trials [F(3,71)=2.60, *p*=0.059, Partial Eta Squared=0.103] (see Table 5 for means and SEs). Post-hoc Bonferroni-corrected pairwise comparisons revealed that within-neighborhood pointing was marginally slower in Experiment 1 compared to Experiment 4 (p=.051); no other pairwise differences were significant (Experiment 1 vs. Experiment 2: *p*=1.00; Experiment 1 vs. Experiment 3: *p*=0.626; Experiment 2 vs. Experiment 3: *p*=1.00; Experiment 2 vs. Experiment 4: *p*=0.704; Experiment 3 vs. Experiment 4: *p*=1.00;).

Pointing Errors

To investigate the possibility that the spatial learning problem was more difficult in Experiment 4 relative to the other experiments, resulting in larger errors on within-neighborhood trials in experiment 4, a cross-experimental comparison was made using a two-way repeated measures Location x Experiment ANOVA of the pointing errors. This analysis revealed main effects of Location [F(1, 68)=86.687, *p*<0.001, Partial Eta Squared=0.56] and Experiment [F(3, 68)=3.988, *p*=0.011, Partial Eta Squared=0.15] and a significant interaction between Location and Experiment [F(3, 68)=6.413, *p*=0.001, Partial Eta Squared=0.221], see Figure 14. Post hoc comparisons showed that pointing errors were larger for between-neighborhood PDO locations (mean=29.30, SE=1.43) than for within-neighborhood PDO locations (mean=42.25, SE=1.43) across experiments, and were significantly larger in Experiment 2-Whole condition than those in

Experiment 1 (*p*=0.019) and Experiment 3 (*p*=0.04) across locations, but no other differences were significant (Bonferroni-corrected) (Experiment 1: mean=31.11, SE=2.31; Experiment 2-Whole condition: mean=42.64, SE=2.98; Experiment 3: mean=32.08, SE=2.31; Experiment 4: mean=37.27, SE=2.31). Separate one-way ANOVAs of within-neighborhood and between-neighborhood pointing errors were used to investigate the significant interaction between Location and Experiment, which revealed a main effect of Experiment for both within- [$F(3,71)=6.14$, $p=0.001$, Partial Eta Squared=0.213] and between-neighborhood trials [$F(3,71)=2.987$, $p=0.037$, Partial Eta Squared=0.116], see Table 5 for means and SEs. Post hoc comparisons showed that pointing errors on within-neighborhood trials were significantly smaller in Experiment 1 than those in Experiment 4 (*p*=0.032) and smaller in Experiment 3 than those in Experiments 2-Whole condition (*p*=0.021) and 4 (*p*=0.005). Pointing errors on between-neighborhood trials were smaller in Experiment 1 than those in Experiment 2-Whole condition (*p*=0.035). There were no other significant differences. (Bonferroni correction was applied).

Discussion

Cross-experimental comparisons revealed that within-neighborhood pointing latencies did not differ between experiments 1 vs. 2 or between 1 vs. 3, undermining the alternative explanation that the "location effect" in experiment 1 (i.e., a within- versus between-neighborhood difference) and total absence of a location effect in Experiments 2 and 3 was simply due to differences in task difficulty. On the other hand, in Experiment 4, response latencies on within-neighborhood pointing trials were marginally longer than those in Experiment 1, while there was no such cross-experiment difference on between-neighborhood trials. This is most likely due to the fact that in Experiment 4, both types of PDOs (shops and restaurants) were inter-mixed throughout the town, eliminating the salient features that might differentiate the two neighborhoods, thereby making within-neighborhood pointing trials just as difficult as between-neighborhood trials.

Cross-experiment comparisons of pointing errors presented a more complex picture. For within-neighborhood pointing trials, errors were smaller in Experiment 1 than 4, smaller in Experiment 3 than 2 (whole condition), and smaller in Experiment 3 than 4. Overall, the cross-experiment comparisons of both pointing latencies and pointing errors suggest that the learning problem faced by participants in Experiment 4 was indeed more difficult than that of the other experiments. This is not surprising, considering that participants were not given any cues that could serve to cluster the environment into smaller regions.

General Discussion

We hypothesized that when people form spatial representations of large-scale complex environments, they may employ either 1) global, flat representations, 2) multi-scale hierarchical representations or 3) independent, local fragments. Each of these makes distinct predictions about accuracy and latency for making within- versus between-region judgements (see Tables 1 and 4). Across the four experiments we found evidence for all three types of representations.

One of our main conclusions is that when participants had fewer trials over which to learn the town layout (Experiment 1), they appeared to form separate, independent representations of the two neighborhoods, whereas when given more learning time (Experiments 2-4), they appeared to form integrated representations of the town. This conclusion is supported by the significant difference in latency when participants were pointing to locations within- versus between-neighborhoods in Experiment 1. Importantly, this effect disappeared when participants were given more time to explore the two neighborhoods jointly (Experiment 2-Whole condition and Experiments 3 and 4). Another interpretation that would be equally consistent with this pattern of results is that in Experiment 1, participants formed hierarchical representations of the two neighborhoods, but the landmarks were only represented at the local, fine-scale level and not at the coarser scale. Thus, when required to point to a landmark in another neighborhood, they would have to integrate their separate, fine-scale representations using post-retrieval processes. In either case, the transition to more global, integrated representations may be a natural progression in the formation of spatial representations of large-scape spaces, as experience in an environment accumulates. Future work could explore this possibility by varying the amount of pre-exposure to individual neighborhoods versus joint exploration of the two neighborhoods.

Another major conclusion from our experiments is that people have a strong tendency to form hierarchical representations. Thus in Experiments 2 and 3, although participants' response latencies were no different for pointing within-versus between-neighborhoods, they were much more accurate on within- compared to between-neighborhood pointing trials. On the other hand, when we removed all visible distinctions between the two neighborhoods in Experiment 4, there was no such effect of location on pointing accuracy. According to our original hypotheses laid out in Table 1, this pattern of findings is consistent with the use of hierarchical representations in Experiments 2 and 3 and a single flat representation in Experiment 4. One potential difficulty with this interpretation is that in Experiment 4 the notion of a "neighborhood" is ill-defined. Our division of the town into two neighborhoods in Experiment 4, using the town midline, was done for convenience, to allow cross-experiment comparisons. From the participants' perspective, there was no obvious basis on which to segregate the town into two neighborhoods, so the entire town would have seemed like a single large neighborhood. Thus our definition of "within-neighborhood pointing" for analysis purposes may have been at odds with participants' perception of what constituted a neighborhood in the town. Further research is required to disentangle the relative contributions of spatially localized distinguishing features versus the size of a region to the formation of spatial representations.

Overall, our results suggest that people have a very strong tendency to form hierarchical representations, grouping space into local regions based on common visual, semantic and geometric features. Early in the acquisition process, people may first form local independent maps of regions, or a less detailed hierarchical representations where landmarks are only represented at fine scales (as suggested by Experiment 1), incurring a reaction time cost when inferring directions between neighborhoods. Once sufficient learning has occurred (Experiments 2, 3), people may form global representations at a coarser scale, in addition to fine-scale regional representations. Only when we removed all available cues that could serve to locally segregate the town into two neighborhoods did people behave as if they formed a single, flat representation of the town, with equal accuracy and latency for within- versus between-region judgements.

The setup of the virtual towns used here was inspired by similar studies in the rodent literature, in which place cells are recorded in environments consisting of two connected boxes. It is interesting to compare these place cell data to the behavioural findings in our own study. In one such study, when rats explored two identical boxes with a connecting corridor, after removal of the barrier in the corridor, place cells underwent a partial remapping, such that some of the CA1 pyramidal cells showed similarly shaped spatial firing fields in both boxes, but others showed completely different spatial firing fields in the two boxes (Skaggs & McNaughton, 1998), suggesting that the rats may have treated the two boxes as separate environments.

Initially we analyzed sex differences, but failed to find any significant differences between males and females in any of the spatial memory measurements. Therefore, we dropped this factor from our current analyses. However, there were other individual differences apparent in our results in terms of sense of direction and GPS usage. Interestingly, GPS usage and video game experience have different effects on spatial memory and recall memory. When the two neighborhoods were distinct, video game players showed better recall memory during the mapping task than non-players. Unfortunately, we do not ask more detailed questions about the type of video game most often played by gamers. It would be interesting to explore, in future research, whether the use of for example first-person-perspective vs aerial perspective is associated with different spatial abilities. Interestingly, in the View condition, greater GPS usage predicted poorer accuracy in making between-neighborhood spatial judgments, suggesting that individuals who frequently rely on a GPS may have greater difficulty integrating their knowledge over large-scale spaces. As noted earlier, the View condition places the greatest demands on participants compared to the Whole or Teleport conditions, requiring them to integrate knowledge between neighborhoods when they have never directly experienced moving from one neighborhood to the other. It could be that people who rely on a GPS when travelling to new places tend to have poorer mental imagery abilities, and thus greater difficulty imagining how the two neighborhoods are connected in the View condition.

What types of spatial representations did participants use in our tasks? Tolman (1948) proposed that two different types of map may underlie spatial cognition: strip-like maps and relatively broad and comprehensive maps. In a strip-map, an animal’s position and the goal position are connected by a single path, which is less flexible in the face of changes in the environment. In contrast, in a comprehensive map, the animal would be able to behave correctly in the face of changes made in the environment. An updated version of this view is that route memories and cognitive maps are two separate and distinct types of spatial representation, with distinct neural substrates (Hartley, Maguire, Spiers, & Burgess, 2003). While the existence of place cells is suggestive of a comprehensive map, it is unclear how knowledge from multiple place cells, each representing a local region of space, is integrated and used to support navigation to a goal. McNaughton and colleagues (1996) suggested that an abstract mental representation of a two-dimensional environment requires input from place cells and head direction cells, which convey self-motion information for path integration and landmark information, respectively. The landmark information could be used to correct for errors accumulated during path integration, but the path integration system should still work without it (McNaughton et al., 1996). Trullier and Meyer (2000) proposed a computational model of navigation, in which they consider the hippocampus as a "cognitive graph" (a.k.a hetero-associative network). The temporal sequences of visited places are learned and an environment's topological representation is formed by using a "place-recognition-triggered response" strategy, which is stored by the network. In this case, the model could predict the next position based on the current position by using place cells, goal cells, and sub-goal cells. Burgess and colleagues proposed a model whereby Hebbian modification of synaptic strengths between place cell representations generates representations of goal locations (Burgess, Jackson, Hartley, & O'Keefe, 2000). Foster, Morris, and Dayan (2000) proposed a hybrid reinforcement learning model that includes a module for learning goal-independent coordinate representations of space and a second module that gradually learns goal-oriented state-action sequences. All of these models suggest ways in which local information from place cells could be integrated into spatial representations.

Place cells appear to be the basis for representing very small local patches of space, and must be integrated to form regional spatial representations. At a larger scale, local spatial representations must somehow be integrated to form global maps of space. Local representations could be connected hierarchically into a multi-scale global representation. Alternatively, they could be chained together via an associative learning process as particular trajectories are experienced, resulting in a representation of a single connected pathway. It is possible that when the two environments are always connected by a path, people may combine them as a chain of local representations. In contrast, when there are multiple connections between the two environments or after extensive experience with the single connection, people may be able to combine them into a global representation. These are important avenues for future research.

Acknowledgements

This research was funded by a grant from the Natural Sciences and Engineering Council to Suzanna Becker and by a scholarship PGS D3 from the Natural Sciences and Engineering Council to Xue Han.

References

Anderson, M. I., & Jeffery, K. J. (2003). Heterogeneous modulation of place cell firing by changes in context. *The Journal of Neuroscience, 23(26),* 8827-8835.

Barry, C., Hayman, R., Burgess, N., & Jeffery, K. J. (2007). Experience-dependent rescalling of entorhinal grids. *Nature Neuroscience, 10(6),* 682-684.

Brockmole, J. R., & Wang, R. F. (2002). Switching between environmental representations in memory. *Cognition, 83,* 295-316.

Burgess, N., Jackson, A., Hartley, T., & O'Keefe, J. (2000). Predictions derived from modeling the hippocampal role in navigation. *Biological Cybernetics, 83,* 301-312.

Byrne, R. W. (1979). Memory for urban geography. *Quarterly Journal of Experimental Psychology, 31(1),* 147-154.

Derdikman, D., Whitlock, J. R., Tsao, A., Fyhn, M., Hafting, T., Moser, M. B., Moser, E. I. (2009). Fragmentation of grid cell maps in a multicompartment environment. *Nature Neuroscience, 12(10),* 1325-1334.

Ekstrom, A., Kahana, M., Caplan, J., Fields, T., Isham, E., Newman, E., et al. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*, 184–187.

Foster, D. J., Morris, R. G. M., & Dayan, P. (2000). A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus, 10,* 1-16.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M. (2004). Spatial representation in the entorhinal cortex. *Science*, *305*, 1258-1264.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436(7052),* 801–806.

Han, X., Byrne, P., Kahana, M., & Becker, S. (2012). When do objects become landmarks? A

VR study of the effect of task relevance on spatial memory. PLoS ONE 7(5), 1-19.

Hartley, T., Burgess, N., Lever, C., Cacucci, F., O'Keefe, J. (2000). Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus, 10,* 369-379.

Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well worn route and the path less traveled: Distinct neural bases of route following and wayfinding in humans. *Neuron*, *37*, 877-888.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*, 425-447.

Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition, 13,* 208-217.

Huttenlocher, J., Hedges, L. V., Corrigan, B., & Crawford, L., E. (2004). Spatial categories and the estimation of location. *Cognition*, *93,* 75-97.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review, 98(3),* 352-376.

Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, *52*, 93-129.

Janzen, G. & van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience, 7,* 673-677.

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., Moser, E. I., & Moser, M. (2008). Finite scale of spatial representation in the hippocampus. *Science, 321,* 140-143.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *The Journal of Neuroscience, 29(31),* 9771-9777.

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology, 18,* 87-121.

McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., Kudrimoti, H., Qin, Y., Skaggs, W. E., Suster, M., & Weaver, K. L. (1996). Deciphering the hippocampal polyglot: The hippocampus as a path integration system. *The Journal of Experimental Biology, 199,* 173-185.

Meilinger, T. (2008). The network of reference frames theory: A synthesis of graphs and cognitive maps. In: *Spatial Cognition VI: Learning, Reasoning, and Talking about Space,* International Conference Spatial Cognition 2008, Springer, Berlin, Germany, 344-360.

Montello, D. R., & Pick, H. L., Jr. (1993). Integrating knowledge of vertically aligned large-scale spaces. *Environment and Behavior, 25(3),* 457-484.

O'Keefe, J., & Dostrovsky, (1971). The hippocampus as a spatial map. *Brain Research , 34,* 171-175.

Paz-Villagrán, V., Save, E., & Poucet, B. (2004). Independent coding of connected environments by place cells. *European Journal of Neuroscience, 20,* 1379-1390.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M. B., et al. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science, 312(5774),* 758–762.

Skaggs, W. E., & McNaughton, B. L. (1998). Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions. *The Journal of Neuroscience, 18(20),* 8455-8466.

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology, 10*, 422-437.

Tanila, H. (1999). Hippocampal place cells can develop distinct representations of two visually identical environments. *Hippocampus,9*, 235-246.

Tolman, E. C. (1948). Cognitive maps in rats and men. *The Psychological Review, 55(4),* 198-208.

Trullier, O., & Meyer, J. (2000). Animat navigation using a cognitive graph. *Biological Cybernetics, 83*, 271-285.

Worden, R. (1992). Navigation by fragment fitting: A theory of hippocampal function. *Hippocampus, 2(2),* 165-188.

Figure Captions

Figure 1. Towns' layouts (21 by 10 VR units in size) used in Experiments. The grey squares are non-distinctive uniformly textured buildings at locations where the participants are not able to drive into. The black squares are places also where the participants are not able to drive into. The red "store" squares and red numbered squares are the stores that serve as passenger drop-off locations. The two green "Start" squares are the two starting points, locating at the two neighborhoods ("Mike's Restaurant" and "Aaron Chang Gallery"). The smaller yellow squares are the two objects, one in each neighborhood. All the white squares indicate locations along the routes that participants can navigate in the town. (a) Town used in Experiments 1 and 2. (b) Town used in Experiment 3, similar to the Town used in Experiments 1 and 2 except the fences between the two neighborhoods were removed. (c). Town used in Experiment 4.

Figure 2. (a) Starting point A (also a tested viewpoint): "Mike's Restaurant". (b) Starting point B (also a tested viewpoint): "Aaron Chang Gallery".

Figure 3. (a) and (b) are the navigators used in the pointing task. At the tip of the compass pointer (red) an image was shown of the target store for the current trial. It could be moved by moving the pointer. (a) Navigator from "Mike's Restaurant" point of view; (b) Navigator from "Aaron Chang Gallery" point of view.

Figure 4. Pointing latency for pointing to PDO locations within vs. between neighborhoods in Blocks 3-5 in Experiment 1. White bar is for within neighborhoods PDO locations and grey bar is for between neighborhoods PDO locations. The pointing latencies were significantly smaller for PDO locations within the neighborhoods than for those between neighborhoods.

Figure 5. Errors for pointing to PDO locations within vs. between neighborhoods in Blocks 3-5 in Experiment 1. White bar is for within neighborhoods PDO locations and grey bar is for

between neighborhoods PDO locations. The pointing errors were significantly smaller for PDO locations within the neighborhoods than for those between neighborhoods.

Figure 6. Pointing latency for pointing to PDO locations within vs. between in each condition in Experiment 2. White bar is for within neighborhoods PDO locations and grey bar is for between neighborhoods PDO locations. Left side bars are for the View condition, middle bars are for the Teleport condition, and right side bars are for the Whole condition. The pointing latencies were not significantly different between the two types of locations, but were significantly faster in the Whole condition than in the View condition.

Figure 7. Errors for pointing to PDO locations within vs. between neighborhoods in each block in each condition in Experiment 2. White bar is for within neighborhoods PDO locations and grey bar is for between neighborhoods PDO locations. Top row is for the View condition, middle row is for the Teleport condition, and the bottom row is for the Whole condition.

Figure 8. Pointing latency for pointing to PDO locations within vs. between in each conditions in Experiment 3. White bar is for within neighborhoods PDO locations and grey bar is for between neighborhoods PDO locations. The pointing latencies were not different between two types of locations.

Figure 9. Errors for pointing to PDO locations within vs. between neighborhoods in each block in Experiment 3. White bar is for within neighborhoods PDO locations and grey bar is for between neighborhoods PDO locations. The pointing errors were significantly smaller for within-neighborhood PDO locations than for between-neighborhood PDO locations.

Figure 10. Pointing latency for pointing to PDO locations within vs. between neighborhoods in Experiment 4. White bar is for within neighborhoods PDO locations and grey bar is for between neighborhoods PDO locations. There was no difference between the two types of PDO locations.

Figure 11. Errors for pointing to PDO locations within vs. between neighborhoods in Experiment 4. White bar is for within neighborhoods PDO locations and grey bar is for between neighborhoods PDO locations. There was no difference between the two types of PDO locations.

Figure 12. Recognition accuracy for within vs. between neighborhoods PDO locations in each Experiment. Solid line is for within neighborhoods PDO locations and dash line is for between neighborhoods PDO locations. Recognition accuracies were significantly worse in Experiment 1 than those in Experiments 2 (Whole condition) and 3 across locations, but no other differences among Experiments 2, 3, and 4. Recognition accuracy for between-neighborhood PDO locations was the worst in Experiment 1 than those in other experiments.

Figure 13. Pointing latency for within vs. between neighborhoods PDO locations in each Experiment. Solid line is for within neighborhoods PDO locations and dash line is for between neighborhoods PDO locations. Pointing latencies were not different among experiments or locations.

Figure 14. Pointing errors for within vs. between neighborhoods PDO locations in each Experiment. Solid line is for within neighborhoods PDO locations and dash line is for between neighborhoods PDO locations.

Table 1. Predictions of pointing latency and pointing error results by different theories

Cross vs. same region pointing			
	Independent fragments	Hierarchical representation	Flat representation
Pointing Latency	Slower	No difference, assuming objects are represented at all levels	No difference
Pointing Error	Larger	Larger	No difference

Table 2. Correlation r between SBSOD scores and recognition accuracy, pointing latency, and pointing errors in Experiment 2.

	Location	View n=16	Teleport n=17	Whole n=12
Recognition Accuracy	Within	0.102 *p*=0.707	0.201 *p*=0.440	0.119 *p*=0.712
	Between	0.143 *p*=0.598	0.247 *p*=0.338	0.523 *p*=0.081
Pointing Latency	Within	-0.351 *p*=0.183	-0.196 *p*=0.451	-0.572 *p*=0.052
	Between	-0.362 *p*=0.169	-0.083 *p*=0.751	-0.591 *p*=0.043*
Pointing Errors	Within	0.040 *p*=0.883	-0.718** *p*=0.001	0.125 *p*=0.698
	Between	0.093 *p*=0.733	-0.418 *p*=0.733	-0.135 *p*=0.676
Number of store recalled		0.439 *p*=0.089	0.449 *p*=0.071	-0.004 *p*=0.990

Table 3. Correlation r between SBSOD scores and recognition accuracy, pointing latency, and pointing errors in Experiments 3 and 4.

	Location	Experiment 3 n=20	Experiment 4 n=20
Recognition Accuracy	Within	-0.076 *p*=0.749	-0.061 *p*=0.779
	Between	-0.018 *p*=0.940	
Pointing Latency	Within	-0.075 *p*=0.755	-0.035 *p*=0.883
	Between	-0.239 *p*=0.311	
Pointing Errors	Within	0.140 *p*=0.557	0.241 *p*=0.306
	Between	-0.217 *p*=0.358	
Number of store recalled		0.159 *p*=0.504	0.022 *p*=0.928

Table 4. Supporting evidence for each of the theories

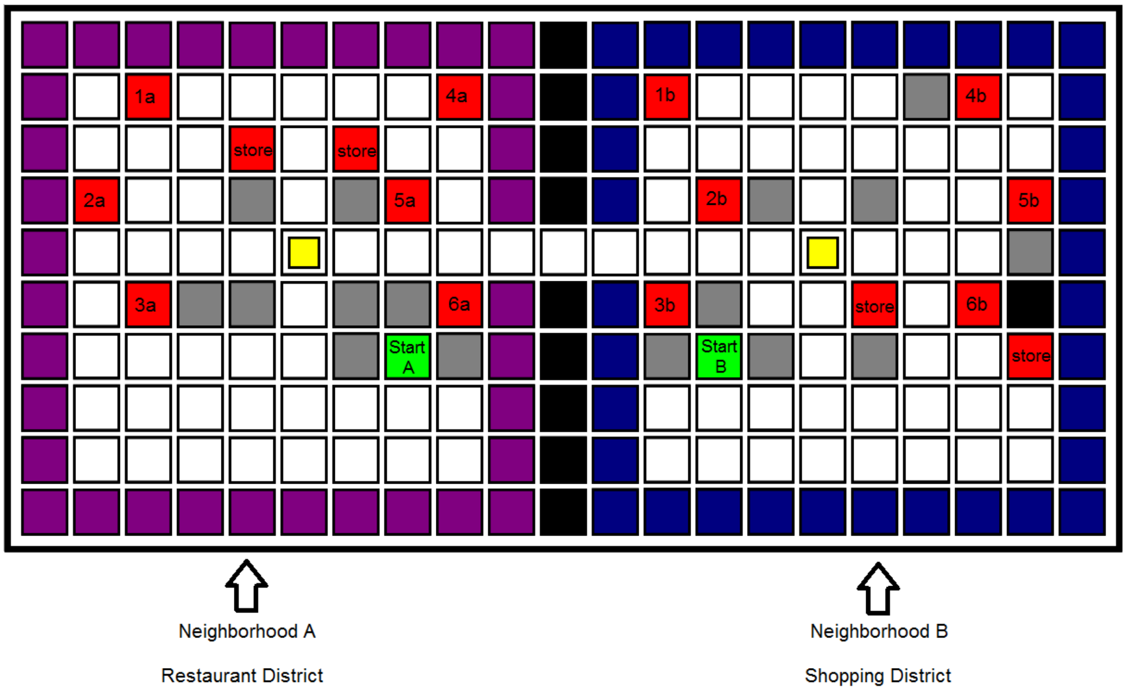
Cross vs. same region pointing			
	Independent fragments	Hierarchical representations	Flat representation
Pointing Latency	Slower	No Diff.	No Diff.
Pointing Error	Larger	Larger	No Diff.
Results	Experiment 1	Experiments 2 & 3	Experiment 4

Table 5. Means and SEs of Recognition Accuracy, Pointing Latency, Pointing Errors in Experiments 1, 2 (Whole condition), 3, and 4.

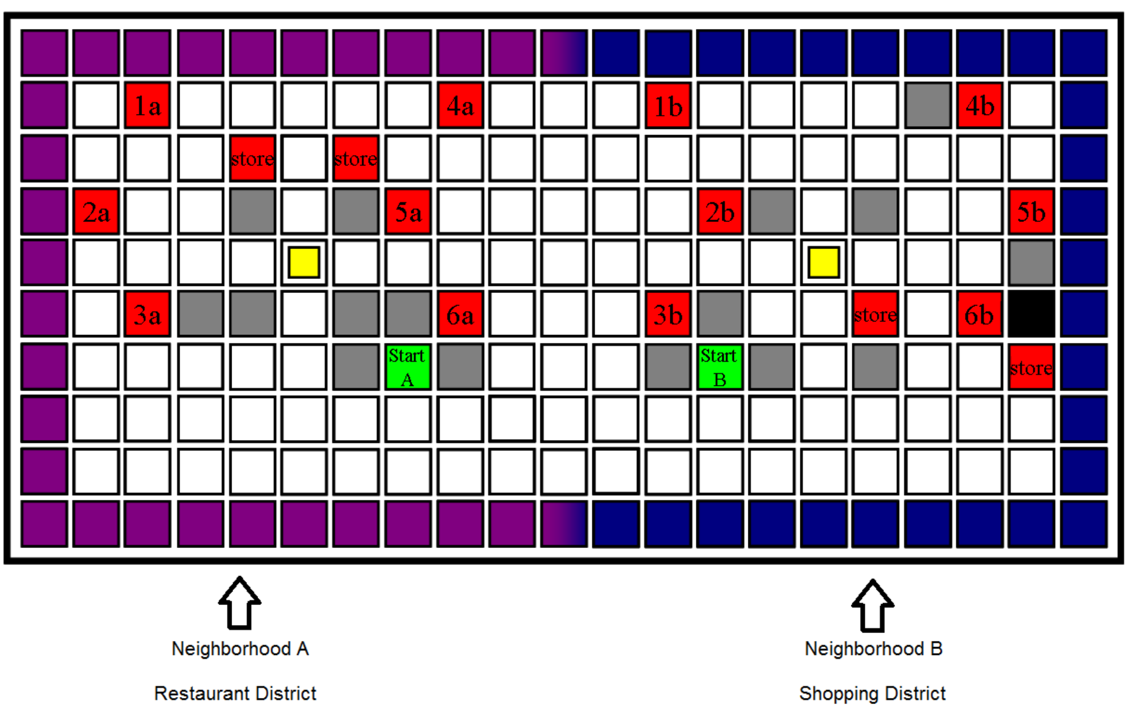
Experiments	N	Recognition Accuracy (in percentage)		Pointing Latency (in seconds)		Pointing Errors (in degrees)	
		Within	Between	Within	Between	Within	Between
Experiment 1	20	95.56% (0.012)	92.78% (0.017)	2.87 (0.24)	3.76 (0.29)	24.49 (1.72)	37.74 (3.57)
Experiment 2 (Whole condition)	12	98.27% (0.006)	97.74% (0.007)	3.21 (0.38)	2.94 (0.34)	35.26 (5.07)	50.01 (3.69)
Experiment 3	20	96.67% (0.008)	98.61% (0.005)	3.50 (0.21)	3.48 (0.24)	22.20 (1.78)	41.95 (1.63)
Experiment 4	20	96.53% (0.008)	97.36% (0.006)	3.91 (0.33)	3.69 (0.30)	35.25 (3.16)	39.29 (2.21)

Figure 1. Towns layouts.

(a) Town used in Experiments 1 and 2.



(b) Town used in Experiment 3.



(c) Town used in Experiment 4

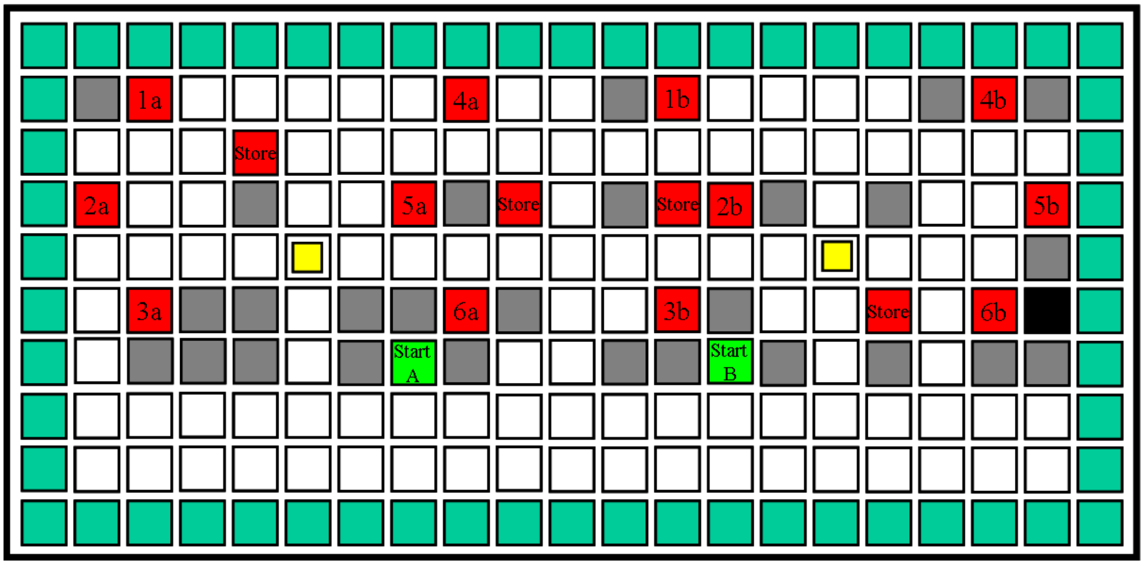


Figure 2. Starting points.

(a) Mike's Restaurant starting point

(b) Aaron Chang Gallery starting point

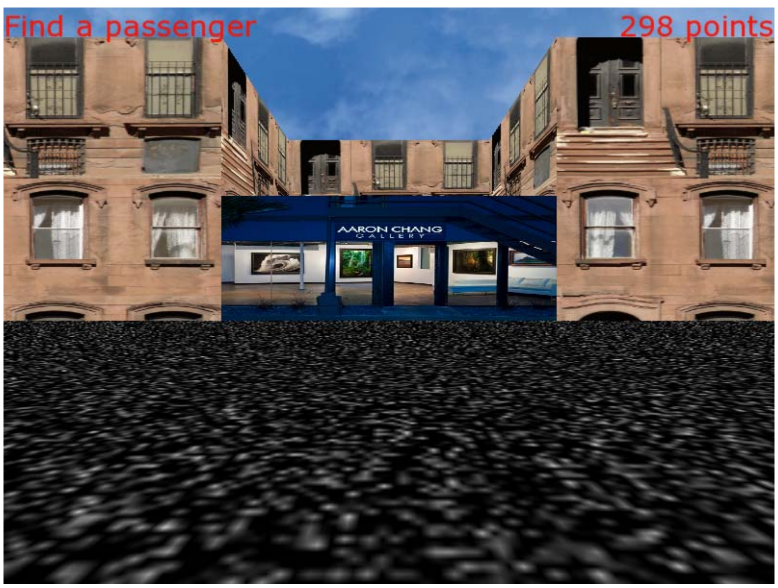
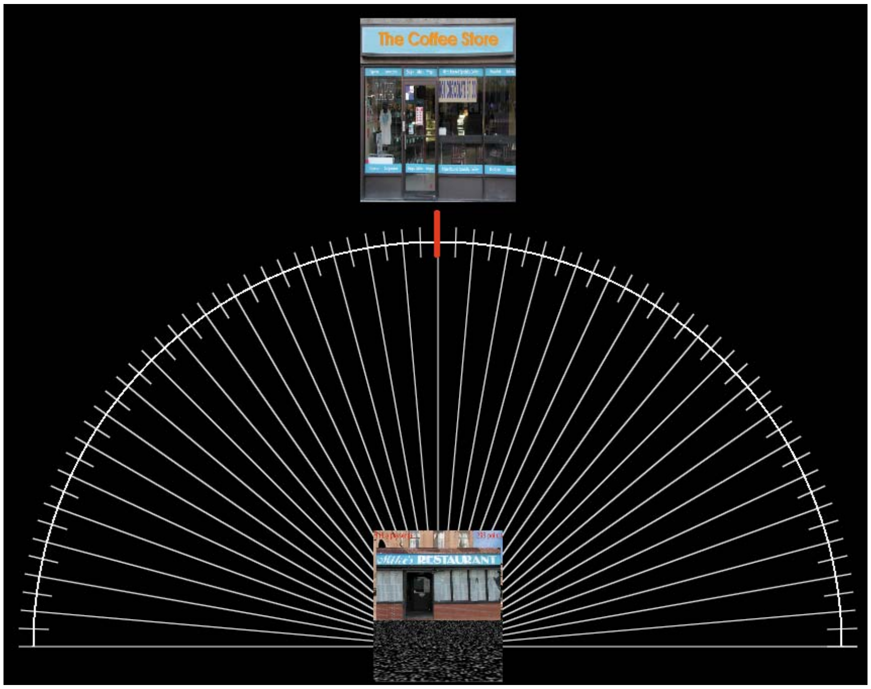


Figure 3. Testing viewpoints.

(a) Tested viewpoint: Mike's Restaurant



(b) Tested viewpoint: Aaron Chang Gallery

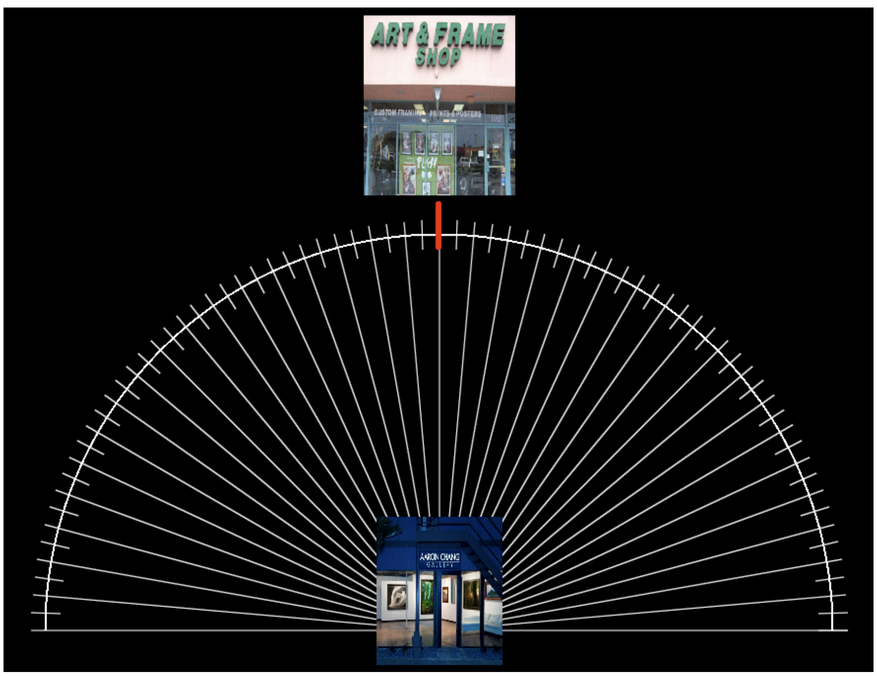


Figure 4. Pointing Latency in Experiment 1

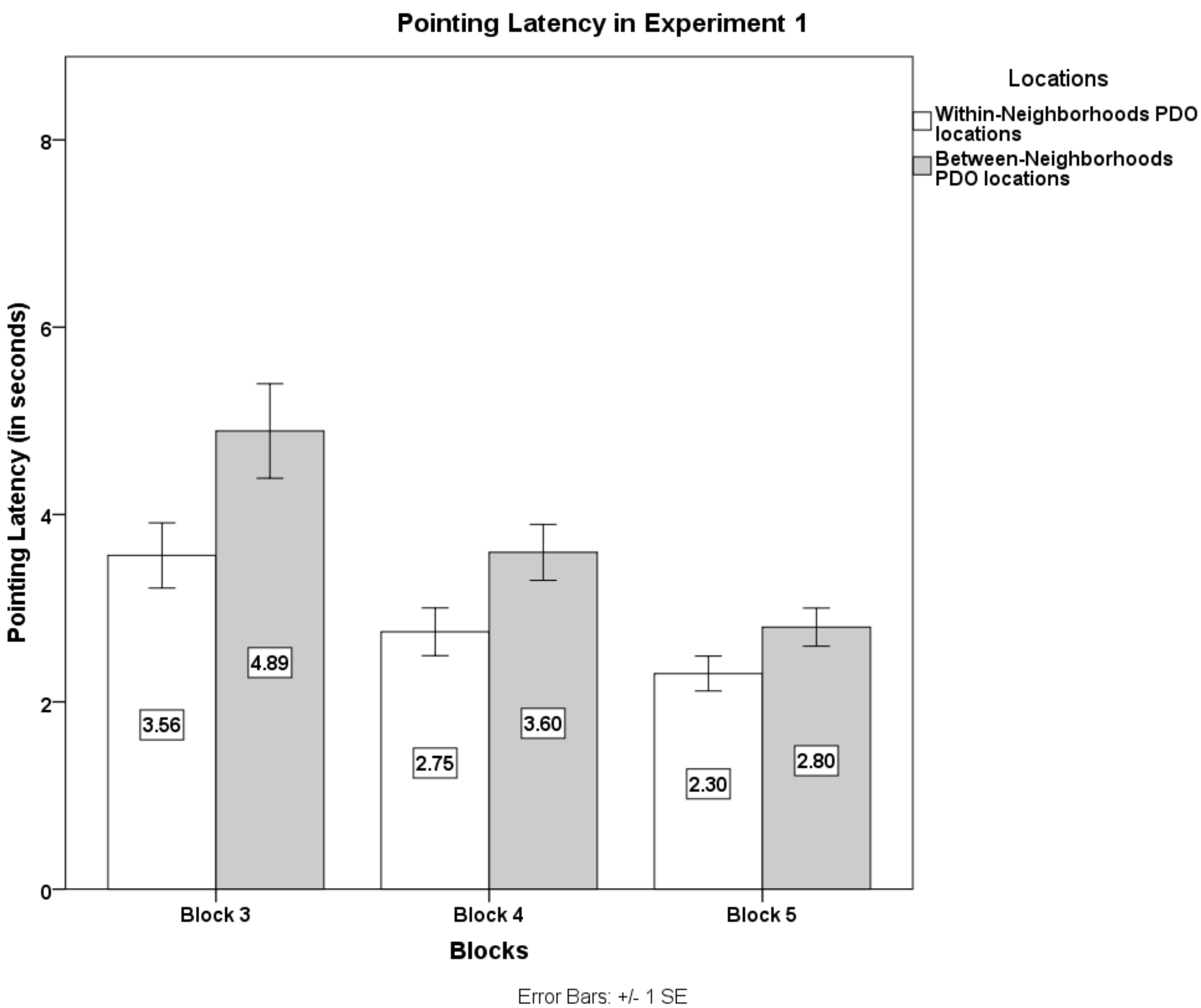


Figure 5. Pointing errors in Experiment 1.

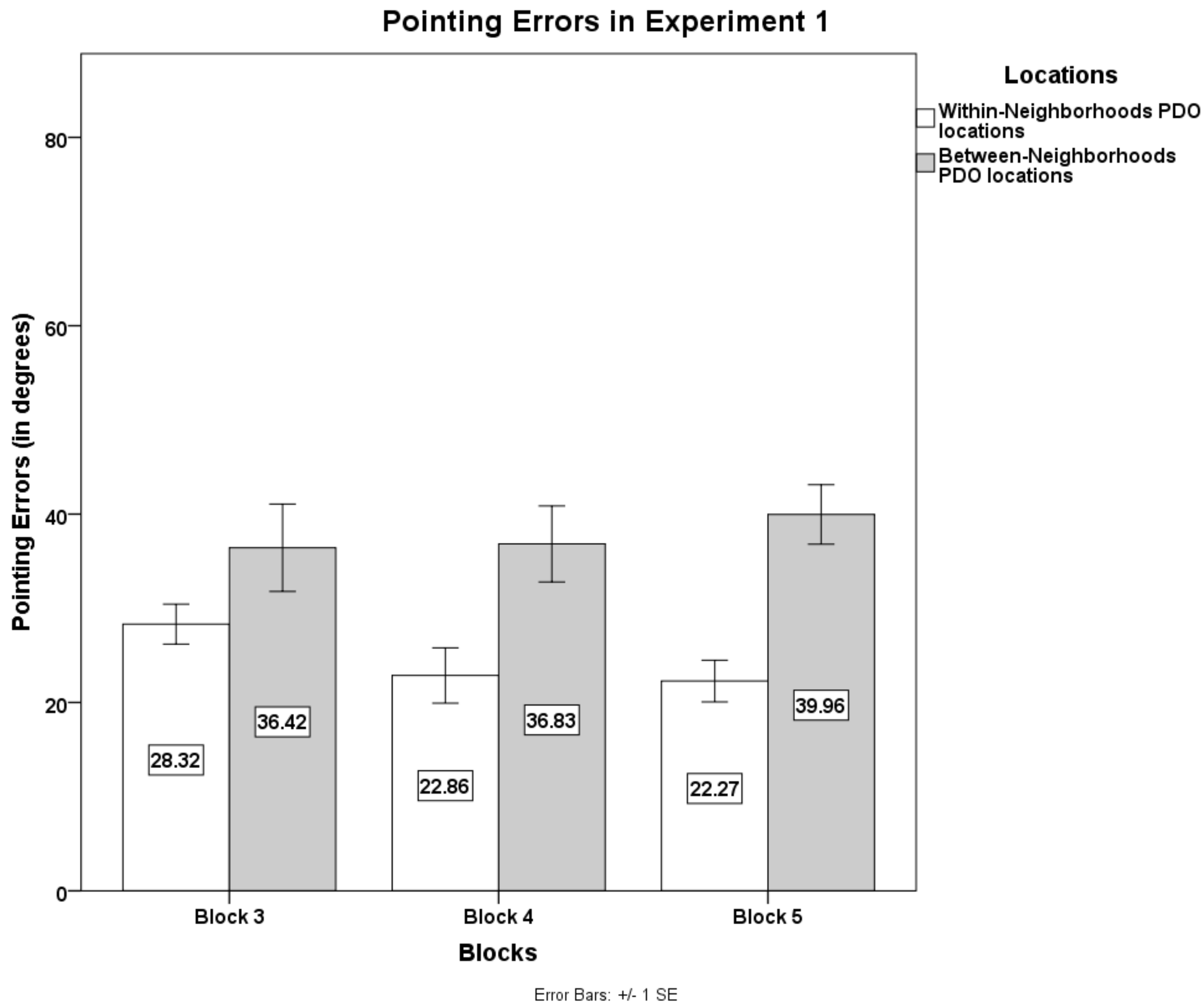


Figure 6. Pointing latency in Experiment 2.

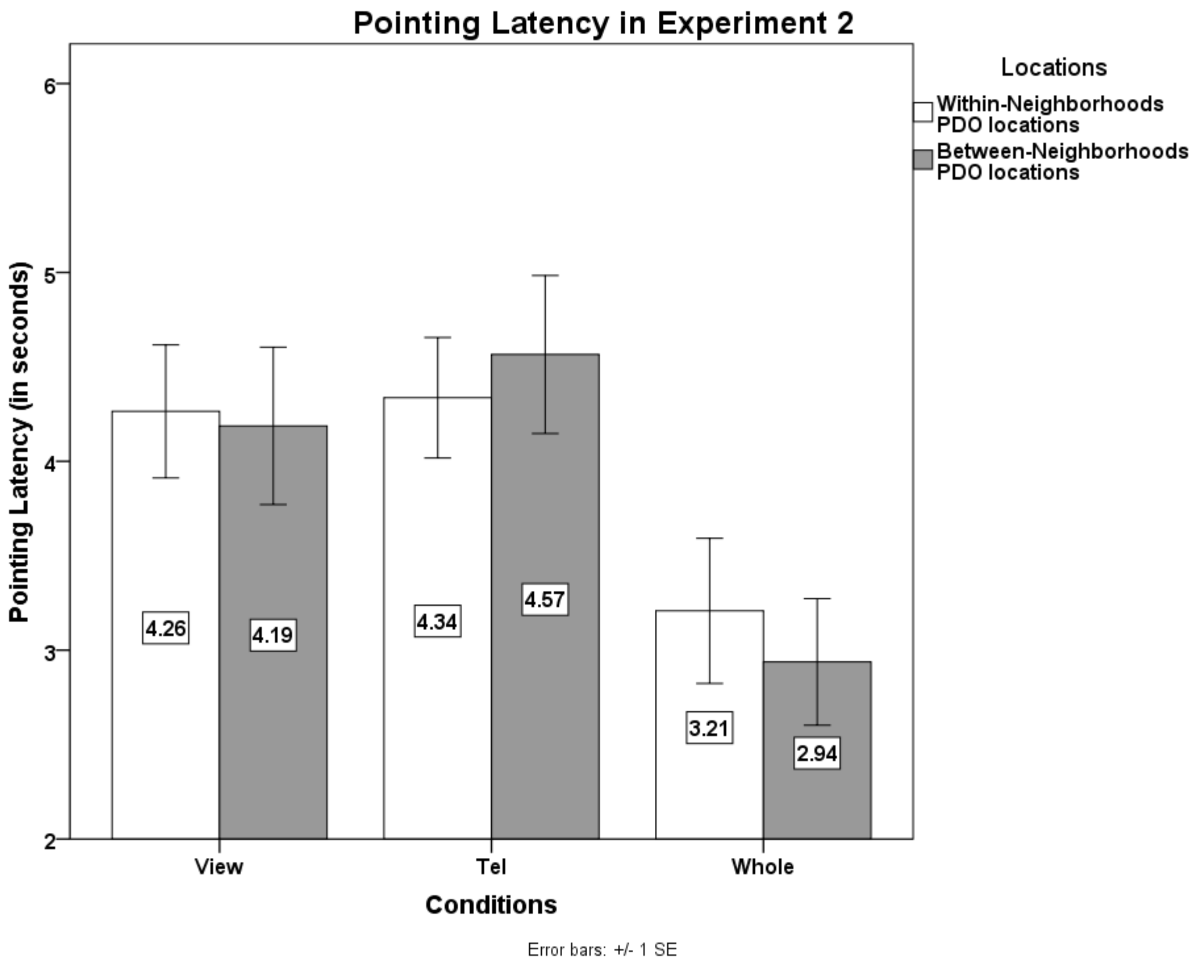


Figure 7. Pointing errors by blocks in Experiment 2.

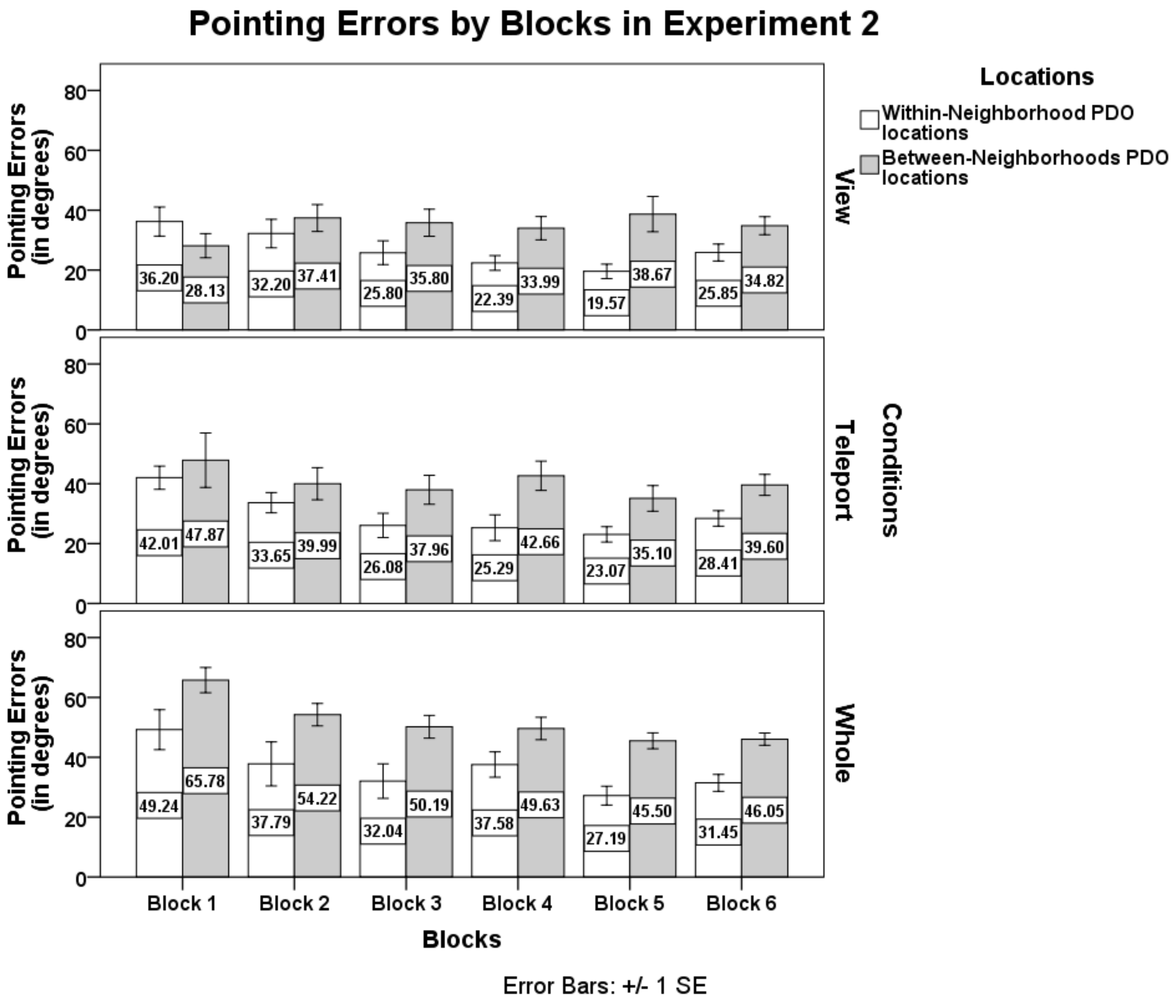


Figure 8. Pointing latency in Experiment 3.

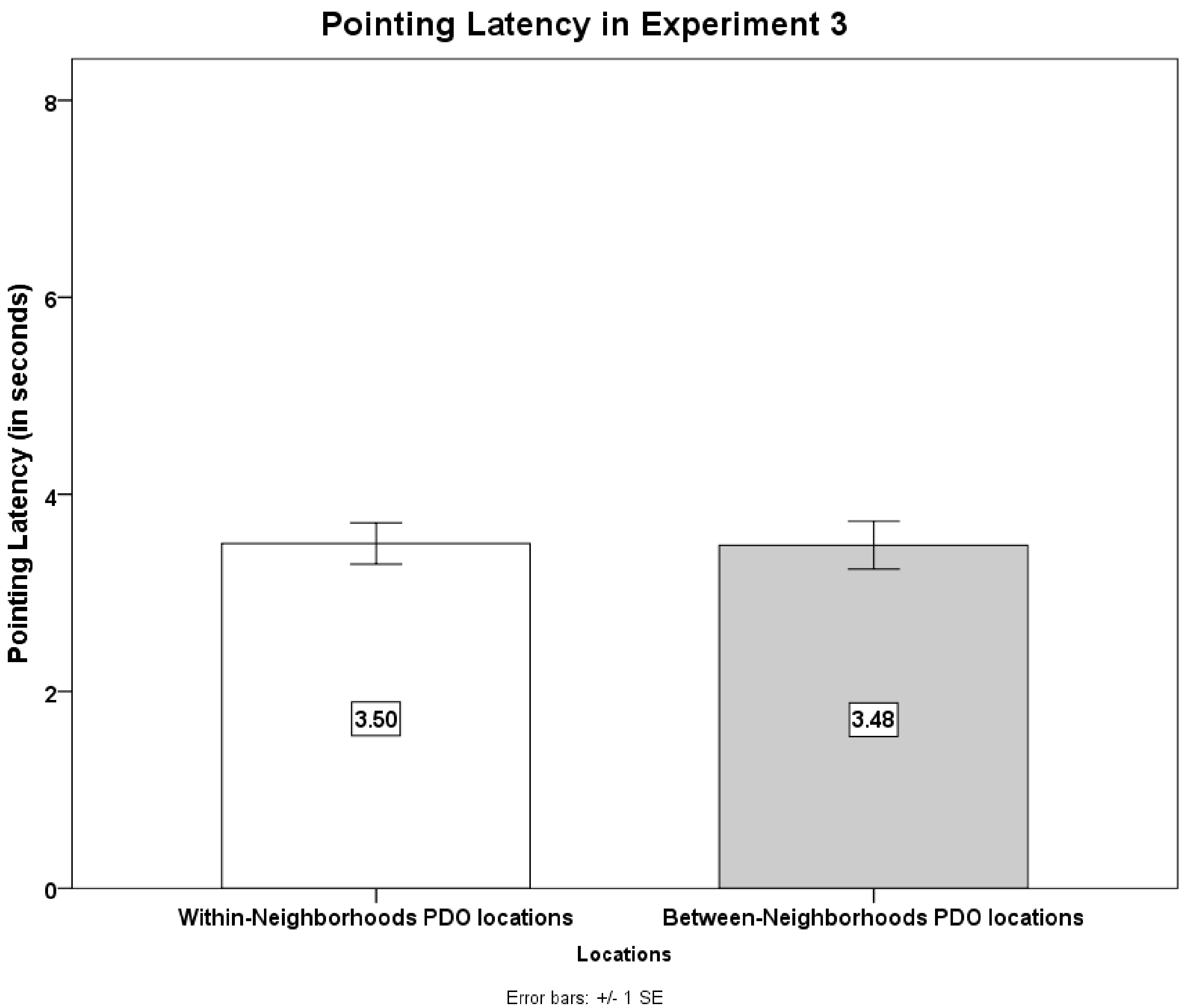


Figure 9. Pointing errors by blocks in Experiment 3.

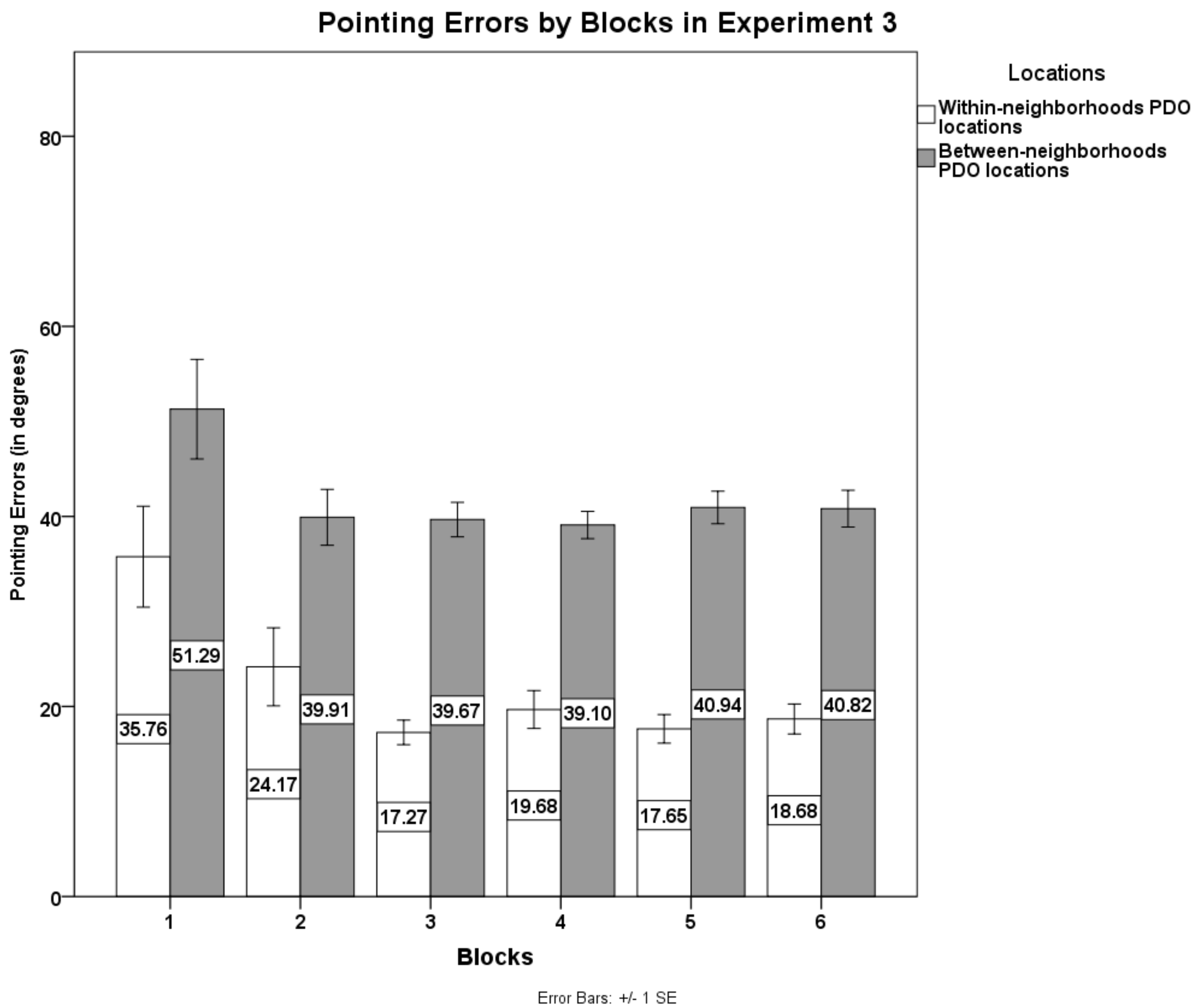


Figure 10. Pointing latency in Experiment 4.

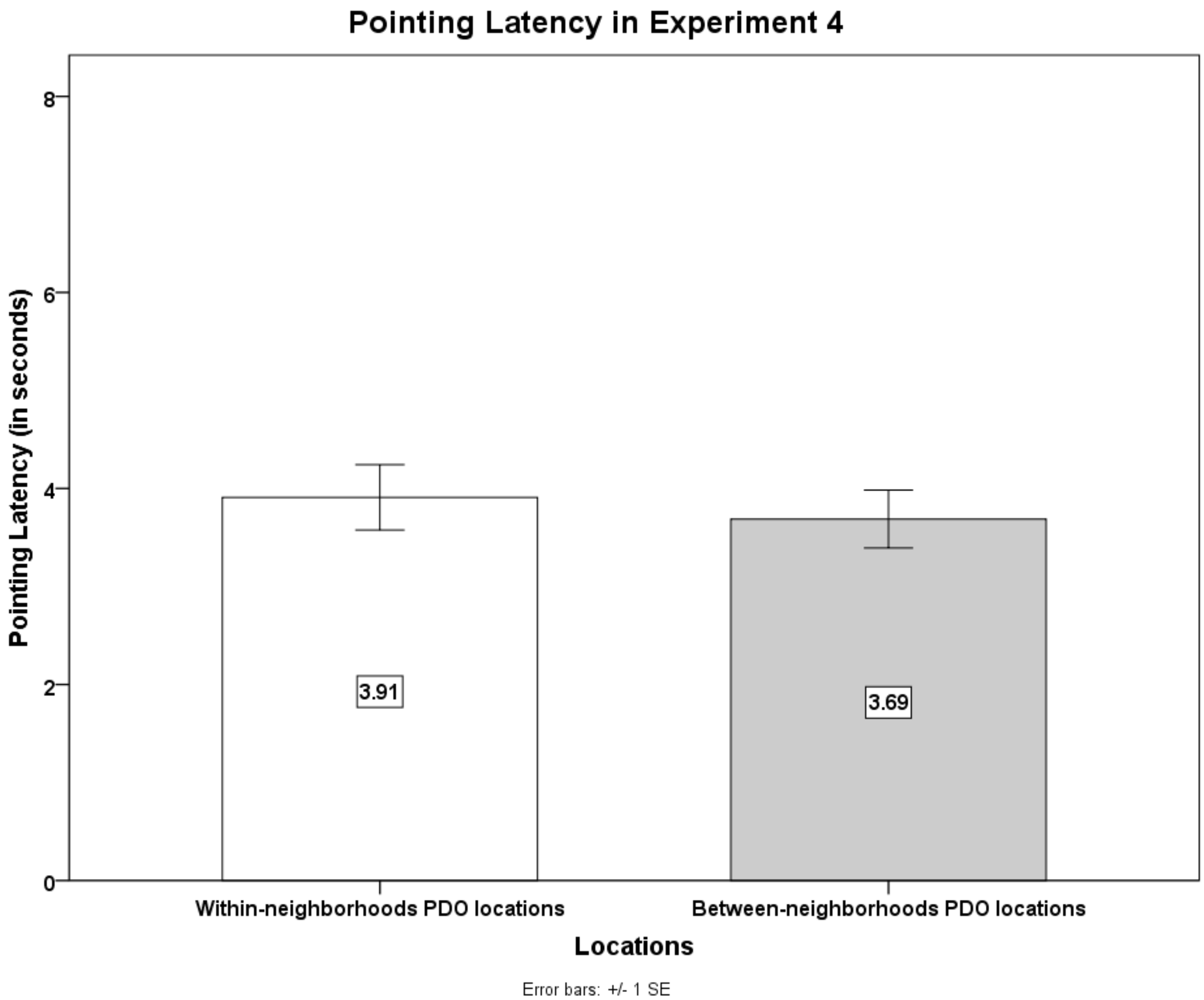


Figure 11. Pointing errors in Experiment 4.

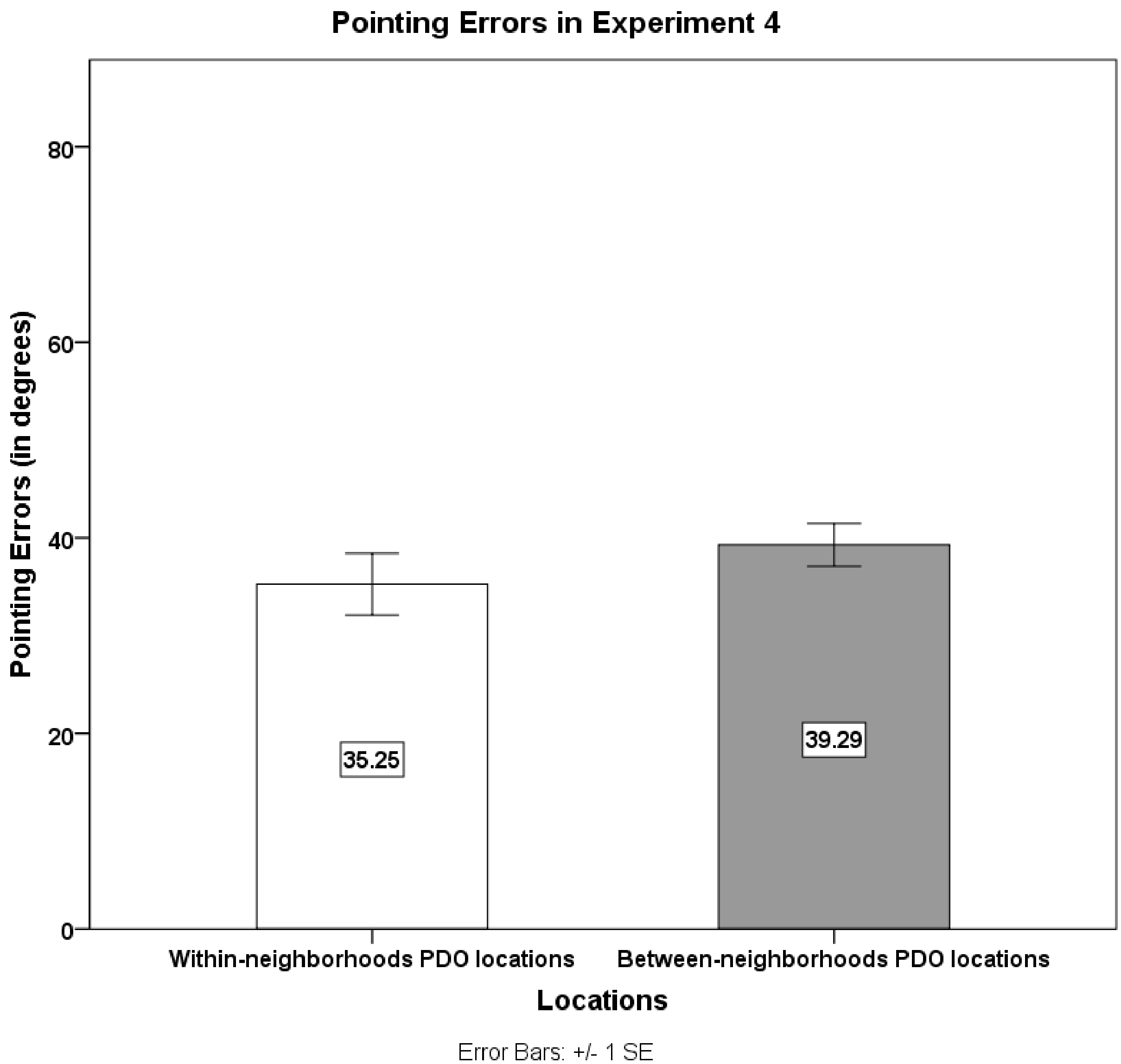


Figure 12. Recognition accuracy in Experiments 1, 2 (Whole Condition), 3, and 4.

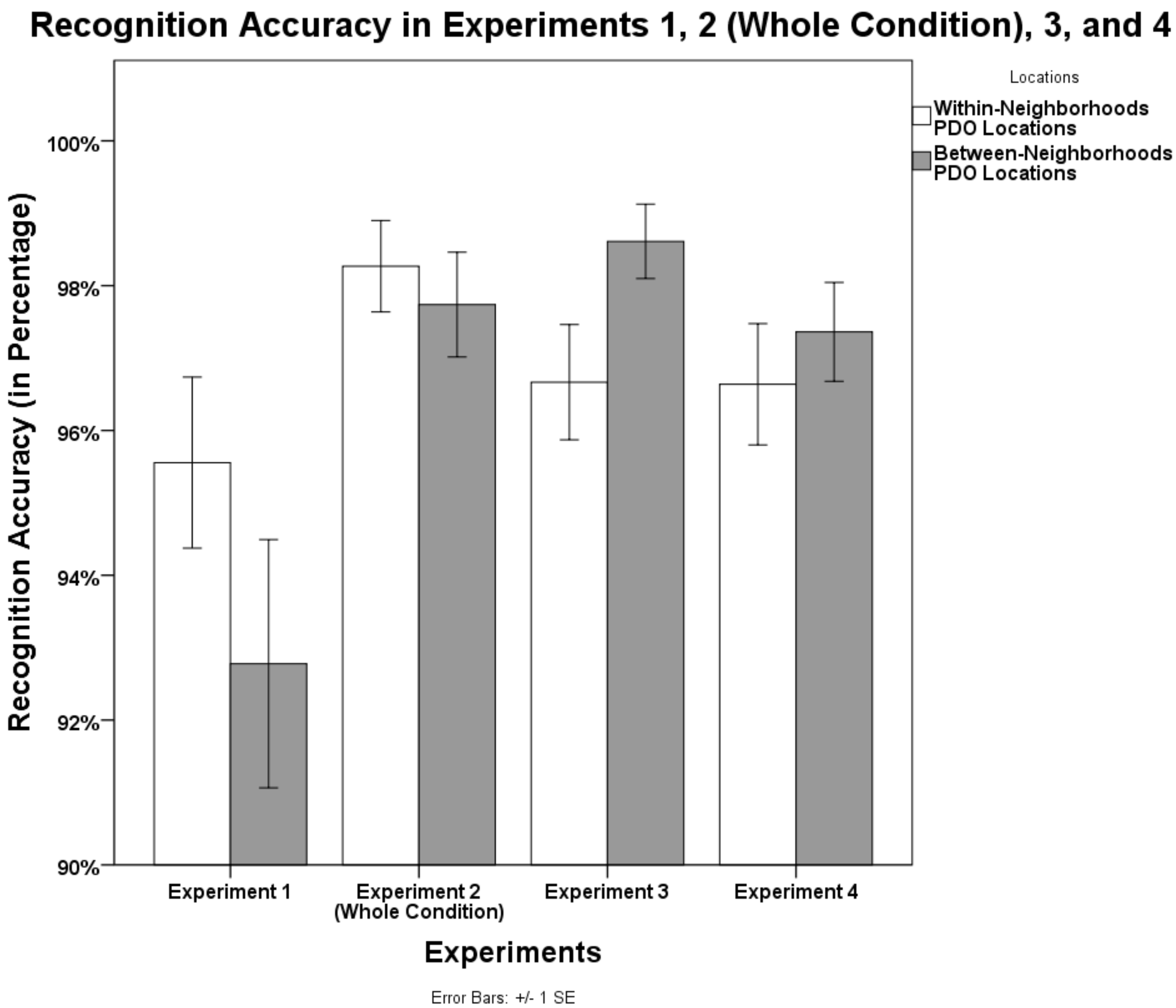


Figure 13. Pointing latency in Experiments 1, 2 (Whole), 3, and 4.

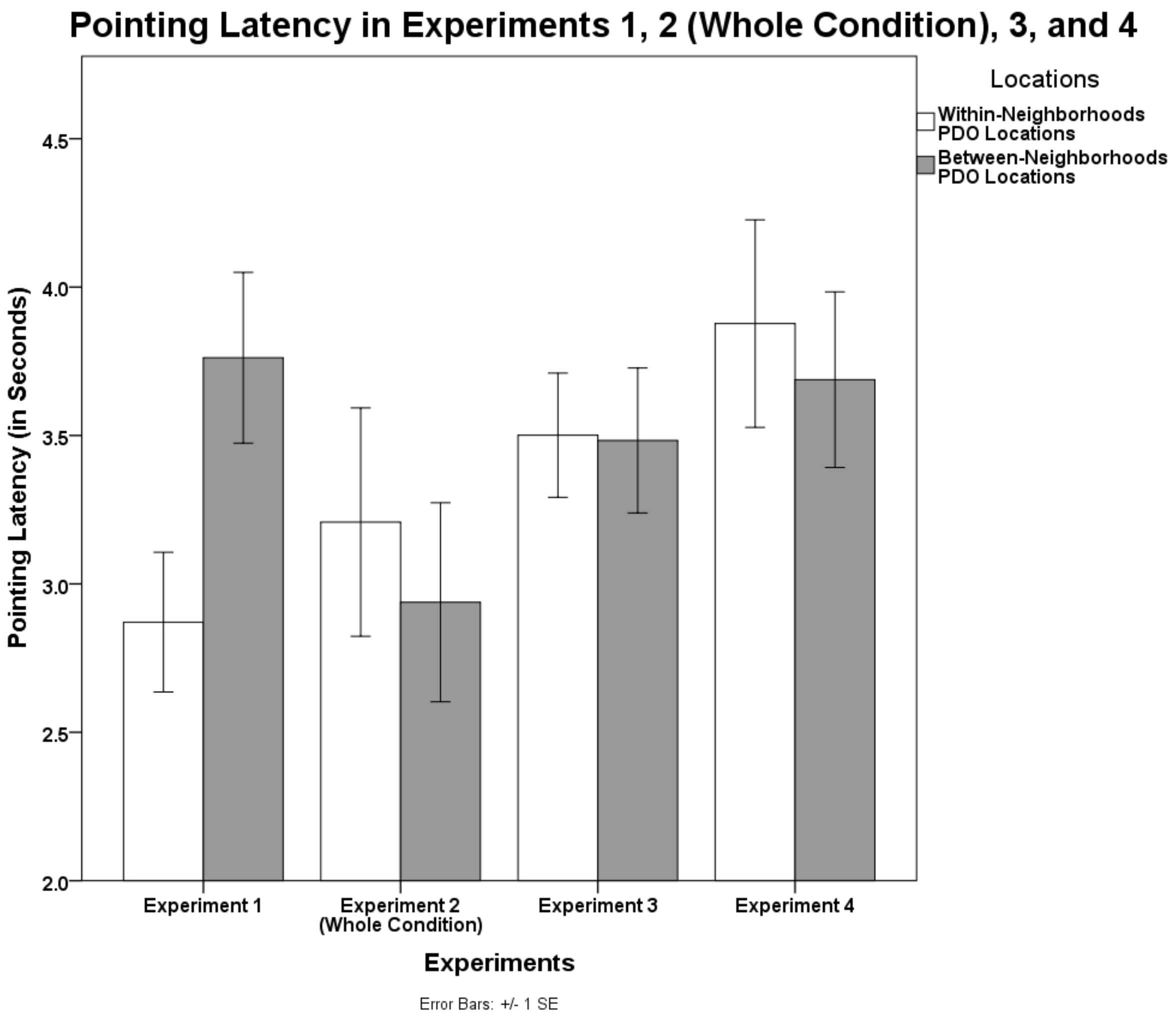
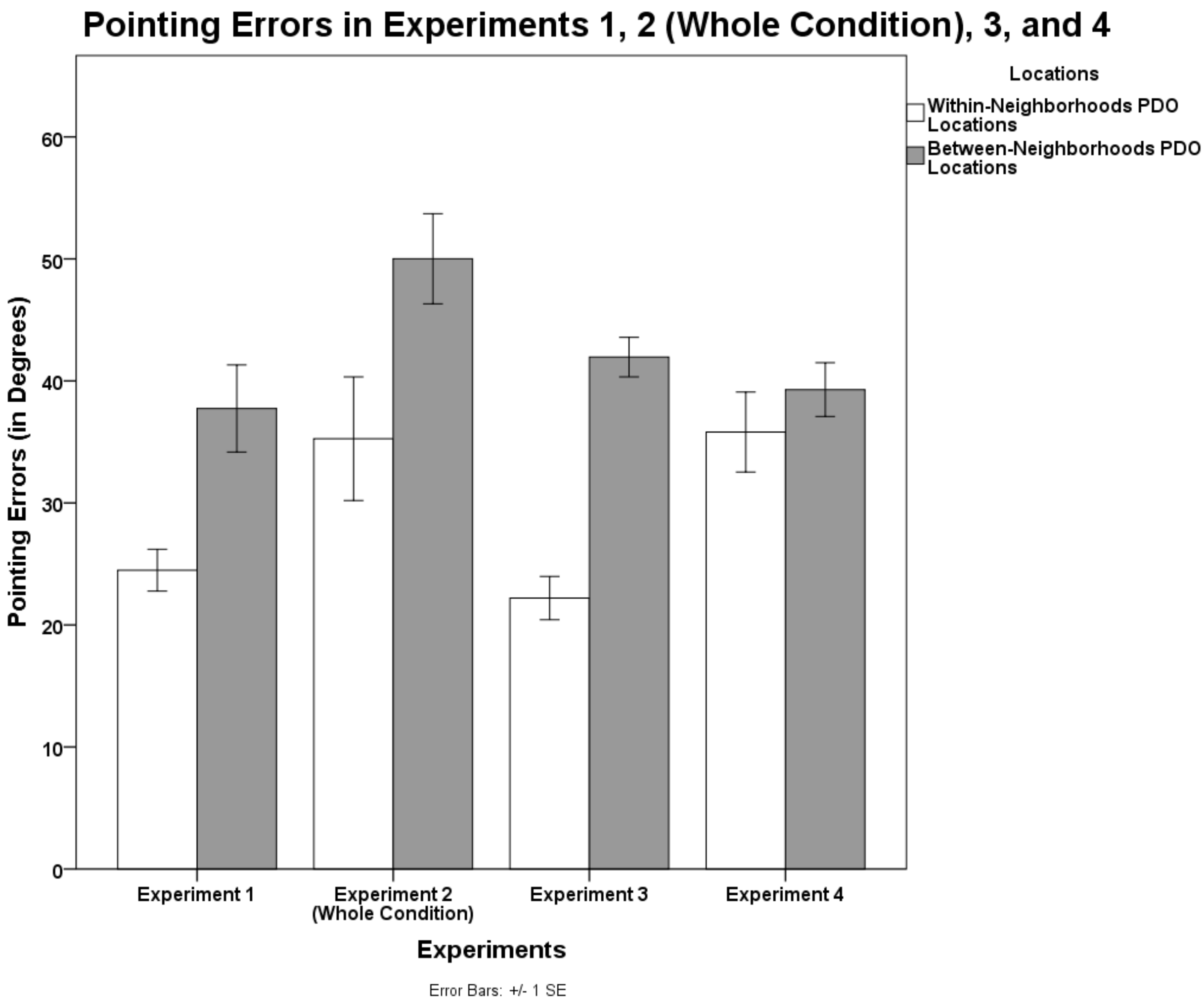


Figure 14. Pointing errors in Experiments 1, 2 (Whole Condition), 3, and 4.



Appendix

Gender: M F Date:

This questionnaire consists of several statements about your spatial and navigational abilities, preferences, and experiences. After each statement, you should circle a number to indicate your level of agreement with the statement. Circle “1” if you strongly agree that the statement applies to you, “7” if you strongly disagree, or some number in between if your agreement is intermediate. Circle “4” if you neither agree nor disagree. For question 17, circle “Yes” or “No” for your answer.

Q1. I am very good at giving directions.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q2. I have a poor memory for where I left things.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q3. I am very good at judging distances.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q4. My “sense of direction” is very good.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q5. I tend to think of my environment in terms of cardinal directions (N, S, E, W).

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q6. I very easily get lost in a new city.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q7. I enjoy reading maps.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q8. I have trouble understanding directions.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q9. I am very good at reading maps.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q10. I don'*t* remember routes very well while riding as a passenger in a car.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q11. I don'*t* enjoy giving directions.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q12. It's not important to me to know where I am.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q13. I usually let someone else do the navigational planning for long trips.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q14. I can usually remember a new route after I have traveled it only once.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q15. I don'*t* have a very good “mental map of my environment.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q16. I use a GPS when I travel to a new place.

(Strongly agree) 1 2 3 4 5 6 7 (strongly disagree)

Q17. I play video games.

a. Yes

b. No