

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

Measuring configural spatial knowledge: Individual differences in correlations between pointing and shortcutting.

Permalink

<https://escholarship.org/uc/item/7t00n522>

Journal

Psychonomic Bulletin & Review, 30(5)

Authors

He, Chuanxiuyue

Boone, Alexander

Hegarty, Mary

Publication Date

2023-10-01

DOI

10.3758/s13423-023-02266-6

Peer reviewed



Measuring configural spatial knowledge: Individual differences in correlations between pointing and shortcutting

Chuanxiuyue He¹ · Alexander P. Boone¹ · Mary Hegarty¹

Accepted: 1 March 2023 / Published online: 17 March 2023
© The Author(s) 2023

Abstract

People use environmental knowledge to maintain a sense of direction in daily life. This knowledge is typically measured by having people point to unseen locations (judgments of relative direction) or navigate efficiently in the environment (shortcutting). Some people can estimate directions precisely, while others point randomly. Similarly, some people take shortcuts not experienced during learning, while others mainly follow learned paths. Notably, few studies have directly tested the correlation between pointing and shortcutting performance. We compared pointing and shortcutting in two experiments, one using desktop virtual reality (VR) ($N = 57$) and one using immersive VR ($N = 48$). Participants learned a new environment by following a fixed route and were then asked to point to unseen locations and navigate to targets by the shortest path. Participants' performance was clustered into two groups using K-means clustering. One (lower ability) group pointed randomly and showed low internal consistency across trials in pointing, but were able to find efficient routes, and their pointing and efficiency scores were not correlated. The others (higher ability) pointed precisely, navigated by efficient routes, and their pointing and efficiency scores were correlated. These results suggest that with the same egocentric learning experience, the correlation between pointing and shortcutting depends on participants' learning ability, and internal consistency and discriminating power of the measures. Inconsistency and limited discriminating power can lead to low correlations and mask factors driving human variation. Psychometric properties, largely under-reported in spatial cognition, can advance our understanding of individual differences and cognitive processes for complex spatial tasks.

Keywords Direction estimation · Wayfinding · Individual differences · Psychometrics

Introduction

Learning the layout of a new environment, that is spatial knowledge acquisition, is a fundamental cognitive function. Humans rely on spatial knowledge to maintain a sense of direction while locomoting through different environments and planning routes to goal locations. Environmental spatial knowledge encompasses different kinds of knowledge, including landmark, route, and configural knowledge (McNamara, 2013; Siegel & White, 1975). Configural knowledge is assumed to integrate all spatial information into a globally consistent mental representation. Compared to landmark and route knowledge, acquiring configural knowledge shows the largest individual differences (Ishikawa & Montello, 2006; Peer et al., 2021; Weisberg

& Newcombe, 2018). It is critical to investigate these individual differences using valid and reliable measures (Newcombe et al., 2023) to advance our understanding of configural knowledge.

Configural knowledge acquisition is typically measured by direction estimation or shortcutting tasks after giving participants a controlled experience of learning routes through a new environment from an egocentric perspective. In direction estimation tasks, participants are asked to point to unseen target locations from different locations and perspectives in the newly learned environment (judgments of relative direction). The fidelity of configural knowledge is measured by average absolute pointing error, that is, the angular disparity between the correct direction and the participant's estimate, averaged across trials (e.g., Ishikawa & Montello, 2006; Meilinger et al., 2014; Schinazi et al., 2013). In shortcutting tasks, participants are asked to take the shortest path to goal locations in the environment, and the measure of performance is wayfinding efficiency, or directness of the path, measured by comparing the path taken to the optimal

✉ Chuanxiuyue He
c_he@ucsb.edu

¹ University of California, Santa Barbara, CA, USA

(shortest) traversable path to the goal location, again averaging over trials (e.g., Gagnon et al., 2016, 2018; Gallistel, 1990; Hartley et al., 2003; He et al., 2019; Tolman, 1948). Note that knowledge of the route that people learn during the learning phase is not sufficient to perform either of these tasks, so they measure how well participants have inferred configural knowledge from the egocentric learning experience. Moreover, in some research paradigms, the walls disappear during wayfinding, so “shortcutting” means straight-line navigation (e.g., Chrastil & Warren, 2013; Foo et al., 2005; Warren et al., 2017). In others, participants cannot go through the walls, and shortcutting means route-based shortcutting (e.g., Chrastil & Warren, 2015; Hartley et al., 2003; He et al., 2021). In the present study, we use the term *shortcutting* to refer to route-based shortcutting.

To examine individual differences in acquiring configural knowledge, researchers have typically used either shortcutting efficiency (e.g., Gallistel, 1990; Hartley et al., 2003) or angular error¹ (e.g., Hegarty et al., 2006; Ishikawa & Montello, 2006; Meilinger et al., 2014; Weisberg et al., 2014; Weisberg & Newcombe, 2018), or have measured pointing and shortcutting performance based on different environments (e.g., Malanchini et al., 2020). Even when both pointing and shortcutting were measured after learning the same environment (e.g., He et al., 2019, 2021; Labate et al., 2014), researchers under-reported the relationship between these measures. It is assumed that they are equally valid and perhaps interchangeable measures of configural knowledge. However, the cognitive demands of estimating the direction to a goal location and of taking the shortest path to that location may not be equivalent. In a route-based shortcutting paradigm, path choices are constrained by the street or path network of an environment (Pagkratidou et al., 2020). In some instances, the shortest path to a goal location may involve temporarily turning away from the direction to the target. Moreover, the ability to point accurately to a goal location is not necessary for efficient wayfinding. For example, participants can take advantage of wormholes to take shortcuts without realizing the physical impossibility of the environment (Murphy & Glennerster, 2018; Warren et al., 2017).

Examining the differential cognitive demands and individual differences in two tasks can thus inform debates on the nature of configural knowledge. One view is that configural knowledge is metrically accurate and globally consistent (Carpenter et al., 2015; Gallistel, 1990; O’Keefe & Nadel,

1978; Siegel & White, 1975; Tolman, 1948), like a physical or cartographic map. Another view is that configural knowledge is labeled graph knowledge, in which close locations are connected with coarse, local metric information (direction and distance) but not metrically consistent across the whole environment (Chrastil & Warren, 2015; Foo et al., 2005; Warren, 2019). Other views are that this distinction is subject to individual differences (Weisberg & Newcombe, 2018) or that map-based knowledge and graph-based knowledge coexist, with the use of different types of knowledge depending on environmental characteristics and navigational demands (Peer et al., 2021). Chrastil and Warren (2015) have proposed that the route-based shortcutting task measures graph-based knowledge and the pointing task measures map-based knowledge.

Here, we examine the correlations between pointing and shortcutting after the same learning experience to address the question of whether they are interchangeable measures of configural knowledge. To address this question, the first step is to examine the psychometric properties of the two measures, as this may affect the correlation between the measures. Based on classical test theory (Novick, 1966; Wilson, 2005), previous researchers have assumed equal difficulty and adequate discriminating power across the items in these measures. The equal difficulty or internal consistency assumption is that participants’ performance on one trial can predict their performance on the other trials. Note that internal consistency is one type of measurement reliability. The adequate discriminating power assumption is that the test items can effectively distinguish people with a high trait level from people with a low trait level. However, the difficulty across trials and discriminating power may vary due to differential availability and saliency of navigation cues such as landmarks and street structure in different trials (Caduff & Timpf, 2008; Röser et al., 2012; Sorrows & Hirtle, 1999), and people may be differentially susceptible to these factors (Andersen et al., 2012; Barhorst-Cates et al., 2021; Coutrot et al., 2022; He et al., 2021; Lawton, 2001; Weisberg & Newcombe, 2016). Ignoring reliability may mislead researchers to conclude a dissociation between the abilities measured by two tasks based on a low correlation, when, in fact, that low correlation is due to the low reliability of the individual measures (Ackerman & Hambrick, 2020; Hedge et al., 2018; Parsons et al., 2019; Newcombe et al., 2023). Ignoring inadequate discriminating power leads to the pitfall that the reported results are only applicable to a subset of the population, whereas others are out of scope due to ceiling or floor effects (Cramer & Howitt, 2005; Kang & MacDonald, 2010; Newcombe et al., 2023).

A secondary goal of the present study was to study the generalizability of our findings across navigation scenarios with and without body-based senses. Previous research has highlighted the importance of body-based internal sensory cues (i.e., proprioception, vestibular system, and motor

¹ Map reconstruction and straight line distance estimation are also common individual differences measures in environmental learning studies (Ishikawa & Montello, 2006; Ruginski et al., 2019; Schinazi et al., 2013; Weisberg et al., 2014), but these two measures are beyond the scope of this paper.

ference) in acquiring map-based configural knowledge. For example, Anastasiou and colleagues (2022) suggested that without body-based cues, people may just acquire graph-based knowledge, whereas, with these cues, and corresponding path integration processes, people gain more precise knowledge including metric distance and direction.

In the present study, we examined the internal consistency and discriminating power of pointing and shortcutting measures after people learned the layout of environments, how these psychometric properties influence correlations between the measures, and the interpretation of these correlations. We also examined psychometric properties and correlations separately for more and less able spatial learners. We conducted two experiments, one in a desktop virtual environment, in which people used a mouse and keyboard to navigate, and one in an ambulatory immersive virtual environment.

The present studies

Method

Participants

Desktop virtual reality study Seventy-two undergraduate students (38 female) participated in this study for course credit. Eight female participants were unable to complete the task due to motion sickness, two were excluded because they failed to reach the target on more than 30% of trials, and five male participants were excluded due to technical issues. Fifty-seven (28 female, median age 19 years, range 18–25 years) were included in the final analysis.

Immersive virtual reality study Fifty-one undergraduate students (27 female) participated in this study for course credit. Three female participants were unable to complete the task due to technical issues or misinterpreting the instructions. Forty-eight participants (24 female, median age 19 years, range 18–25 years) were included in the final analysis.

A statistical power analysis showed that with $N = 48$, we could detect a correlation of .4 (a medium effect size: Cohen, 1988) with $\alpha = .05$ and $\text{power} = 0.80$.

Materials

Desktop virtual reality study

Virtual maze

The 11×11 m experimental maze, as shown in Fig. 1a and b, was taken from Boone et al. (2019) (Maze 1). Twelve landmarks were placed in alcoves in the maze (see Fig. 1a).

During the learning phase, people learned the environment by taking a fixed tour of the maze five times.

The experiment was administered using a Dell XPS with a GeForce GTX 1070 graphics card. The environment was presented using Unity3D and displayed on a 24-in. LCD monitor (289.9×531.4 mm display area), with a refresh rate of 60 Hz at a resolution of $1,920 \times 1,080$ and a viewing distance of approximately 1 m.

Direction estimation task

The direction estimation task was conducted using E-prime 2.0 (Schneider et al., 2012) and was administered twice for each participant, once before the shortcutting task (Pointing Phase I) and once after the shortcutting task (Pointing Phase II). On each trial, participants were shown an image of a landmark (starting landmark) on the left half of the screen. An arrow circle was displayed on the right half of the screen (see Fig. 2a). Participants were instructed to imagine being in the maze and facing the starting landmark and to indicate the direction to another (target) landmark (which was not visible from the current location). For example, in one trial, participants were shown a picture of the chair and were asked to point to the well (see Fig. 2a). They indicated the target landmark by dragging a line (a rotating “pointer”) on the displayed arrow circle. There were 27 trials, and the score on this task was the average angular error across trials (Pointing Error). Twenty of these trials used the same starting and target landmarks as the shortcutting task.²

Shortcutting task

In the shortcutting task, participants were positioned at different locations in the maze and instructed to navigate to target landmarks using the shortest path. There were 20 shortcutting trials, which were presented in random order. The shortest path on each trial was at least 25% and on average 51% shorter than the learned route. Participants had 40 seconds to complete each trial. At the end of each trial (finding the target or timing out), participants were transported to the starting location of the next trial.

Immersive virtual reality study

Virtual maze

The 7×6.5 m experimental maze, as shown in Fig. 1c–d, had a similar structure to the desktop study and the same

² For the analysis only including the corresponding subsets, see the [Online Supplementary Materials \(OSM\)](#). The conclusions do not change.

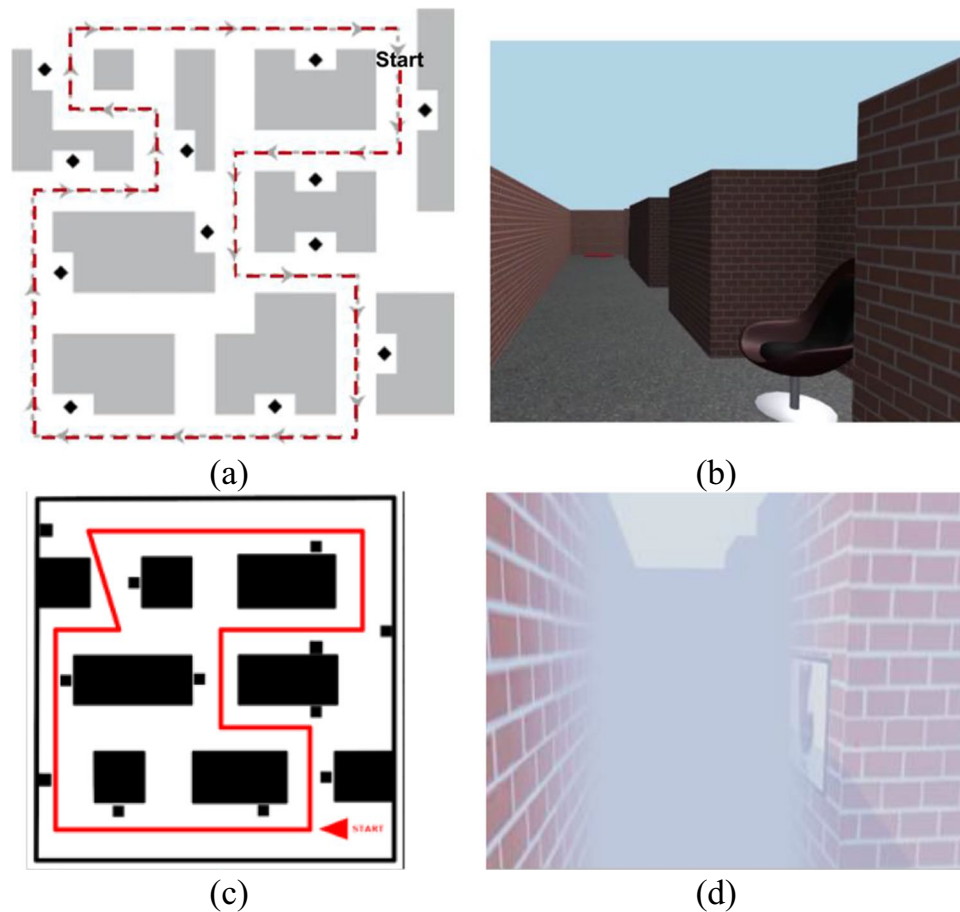


Fig. 1 **a** Map of the virtual environment used by Boone et al. (2019), the red dashed line indicates the route people use to learn the environment during the learning phase. **b** Participants' view of the desk-

top environment. **c** Map of the immersive virtual environment, the red line indicates the route for learning. **d** Participants' view of the immersive virtual environment

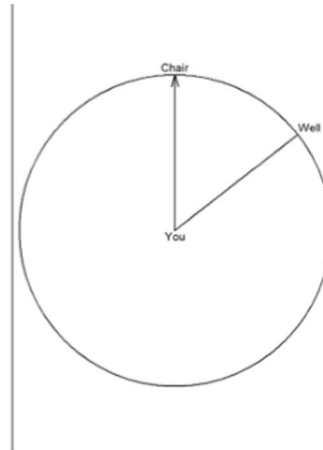
12 landmarks. However, given the physical space constraint of the laboratory, it has a smaller scale and we replaced the 3D objects with pictures of these objects on the walls. Condensing the structure leads to higher visibility, compared to the desktop environment, which means participants can gain more visual information about the structure of the environment at some locations. To make this environment more comparable to the desktop study, we added fog (see Fig. 1d). The fog obscured vision beyond 2.5 m and the clarity decreased linearly between 1 and 2.5 m.

The immersive virtual environment was displayed using an HTC VIVE Pro Eye VR head-mounted display (HMD) with a Dual OLED 3.5-in. diagonal display (1,440 × 1,600 pixels per eye or 2,880 × 1,600 pixels combined), a 90-Hz refresh rate, and a 110° field of view capable of delivering high-resolution audio through removeable headphones. In addition to the HMD, the VR interface included two HTC VIVE wireless handheld controllers for interacting with

the experiment and four HTC Base Station 2.0 infrared tracking sensors for large-scale open space tracking. The system was equipped with wireless room tracking via a 60-GHz WiGig VIVE Wireless adapter and was run on an iBuyPower desktop computer powered by an eight-core, 3.60 GHz Intel core i9-9900K central processing unit (CPU), an NVIDIA GeForce RTX 2070 Super graphics processing unit (GPU) with 16 GB of system memory. Participants physically walked in the environment while wearing the HMD.

Direction estimation task

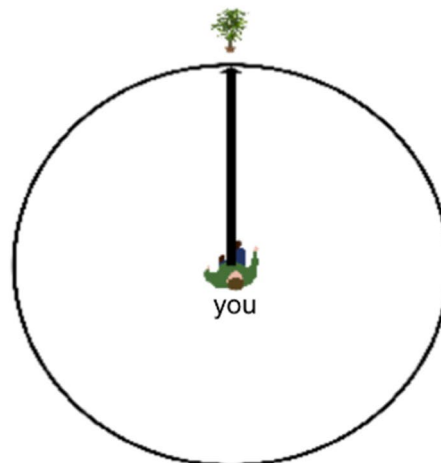
As shown in Fig. 2b, the direction estimation task in the immersive VR study was similar to the desktop study and was run on the desktop, except that the task was programmed in Unity and had 24 trials in total. The 24 trials had the same landmark combinations as the shortcutting task but switched the starting and target landmarks. For



Imagine that you're standing directly in front of the chair and facing the chair, please point to the well

(a)

**Imagine you are standing in front of and facing the plant.
Please point to the bookshelf**



(b)

Fig. 2 **a** The sample item on the instruction screen for the direction estimation task in the desktop study. **b** The sample item for the direction estimation task in the immersive study

example, in the shortcutting task, participants were asked to start from the bookshelf to navigate to the plant, but in the direction estimation task, participants were asked to face the plant and point to the bookshelf. We implemented this change to reduce the impact of the direction estimation task on the shortcutting task. On each trial (as shown in Fig. 2b), participants were instructed to imagine being in the maze and facing the starting landmark, and to indicate the direction to another (target) landmark. They indicated the target landmark by dragging a line (a rotating “pointer”) on the displayed arrow circle (see Fig. 2b). The score on this task was the average angular error across trials (Pointing Error).

Shortcutting task

The shortcutting task was similar to the desktop study except that participants physically walked in the environment and had 24 trials. Participants had 30 seconds for each trial. Between trials, to disorient participants from the previous trial and relocate participants to a new starting location, they were placed in an empty space with floor and visual checkpoints. They were asked to walk to a random checkpoint and then to another checkpoint, placing them in the position and orientation to start a new trial. The 24 trials were selected to ensure the following criteria: (1) each landmark was the

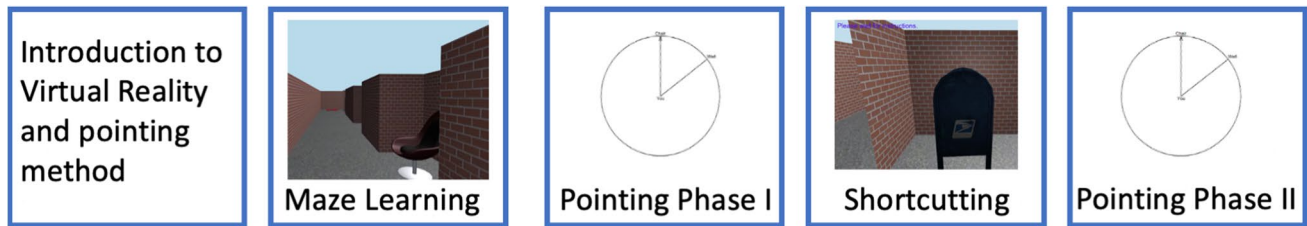


Fig. 3 The order of tasks in the two experiments

start location twice; (2) each landmark was the target at least once but no more than three times, and (3) the shortest path on each trial was at least 30% and on average 49% shorter than the learned route.

Procedure

The local Institutional Review Board (IRB) reviewed and approved both studies as adhering to ethical guidelines. In the desktop study, all participants completed the experiment in a lab cubicle alone, with an experimenter giving instructions. In the immersive study, all participants completed the experiment in the immersive VR lab alone, with one experimenter giving instructions and one experimenter handling the computers. For both studies, after giving informed consent, participants were trained to use the digital arrow circle on the computer screen to indicate directions. Their comprehension of how to indicate directions was checked by having them use the arrow circle to point to two visible objects in the experiment room.

Participants then practiced using the active navigation controls (Desktop: keyboard and mouse; Immersive: headset and controllers) in a training maze.³ Next, participants were placed in the experiment environment maze with red arrows and followed these arrows to learn a route through the virtual environment five times, saying the name of each object aloud as it came into view the first time. After participants followed this route five times, three spatial tasks were administered in a fixed order: (1) direction estimation task – Phase I, (2) shortcutting task, and (3) direction estimation task – Phase II, see Fig. 3.⁴ Finally, participants were debriefed.

All analyses were carried out using Python scripts.

³ The training mazes had different structure to the mazes used in the tasks and had no landmarks. In the desktop study, participants practiced using mouse and keyboard to follow arrows along the floor until comfortable. In the immersive study, participants practiced walking to three gray bubbles and using the controllers to click bubbles. They were also given the time to freely explore the training maze until comfortable.

⁴ We also included an onsite direction estimation task (pointing in the environment) for the desktop study after pointing Phase II for exploratory analyses, which are not included in this paper.

Results

Overall performance

Descriptive statistics, including the internal consistency of the measures, are presented in Table 1. Participants were generally successful in reaching the target within the time limit in both the desktop and immersive VR studies, except for one trial in the desktop study in which 17 of the 57 participants (30%) were unsuccessful; this trial was excluded from way-finding analyses. Participants were successful on 92.9% of the remaining trials in the desktop study and on 94.5% of the trials in the immersive study. Travel Efficiency was defined as the ratio of the distance traveled to the distance of the shortest traversable path on each trial. If a participant took the shortest path on every trial, their efficiency would be 1, and if they took the learned path on every trial, their efficiency would be 2.54 on average for the desktop VR maze (i.e., the average learned route efficiency) and 2.19 for the immersive VR maze. Travel efficiency for the unsuccessful trials was replaced by the average learned route efficiency.⁵

As shown in Table 1, the average pointing error (angular error) in Phase I direction estimation was 74.71° ($SD = 23.22$) and 64.58° ($SD = 27.45$), respectively, for the desktop and immersive environments. Although relatively poor, average performance across all participants was significantly better than chance (90°), one-sample $t(56) = -5.30$, $p < 0.001$, $d = -.70$, 95% CI = [67.54, 79.87] in Desktop and one-sample $t(47) = -6.42$, $p < 0.001$, $d = -.93$, 95% CI = [56.61, 72.55] in Immersive.

The average travel efficiency score across trials was 1.81 for the desktop VR environment and 1.56 for the

⁵ Note that using this substituting method, participants get a penalty (or their efficiency was deflated) if they fail to locate a target, because taking the learned route is an inefficient method in the current paradigm. Another method is removing both the unsuccessful shortcutting trials in calculating efficiency and the corresponding trials in the pointing task. The conclusions do not change if we use this alternative method. Detailed results based on the alternative method are shared online (<https://github.com/CarolHeChuanxiuyue/ConfiguralSpatialKnowledgeMeasurement.git>).

Table 1 Descriptive statistics for pointing error and efficiency for all participants

	Study	Mean	SD	Min	Max	Skewness	Kurtosis	# of Trials	Reliability
Pointing Error (Phase I)	Desktop	73.71	23.22	18.85	105.26	-0.69	-0.27	27	0.83
	Immersive	64.58	27.45	8.21	102.97	-0.67	-0.69	24	0.87
Efficiency	Desktop	1.81	0.39	1.00	2.51	-0.32	-0.93	19	0.72
	Immersive	1.56	0.39	1.04	2.58	0.61	-0.48	24	0.87
Pointing Error (Phase II)	Desktop	65.37	23.40	15.89	104.67	-0.25	-0.83	27	0.84
	Immersive	52.90	28.11	12.49	99.97	0.07	-1.46	24	0.90

immersive VR environment. Therefore travel distance was, on average, significantly shorter than the learned route (Desktop: one-sample t test (56) = -14.02, $p < 0.001$, $d = -1.86$, 95% CI = [1.71, 1.91]; Immersive: one-sample t test (47) = -10.99, $p < 0.001$, $d = -1.59$, 95% CI = [1.45, 1.68]). Notably, in the shortcutting trials, most participants took paths that were shorter than the learned route, although their pointing performance was relatively poor. This is illustrated in Fig. 4 in which the horizontal line indicates chance-pointing performance and the vertical red line indicates the efficiency score of a person who always takes the learned route.

The observed and disattenuated correlations between the measures are shown in Table 2. Disattenuated correlations take the internal consistency (i.e., permutation-based split-half estimation)⁶ of the measures into account using Formula (1) (Parsons et al., 2019; Spearman, 1904) where $r_{observed}$ is the observed correlation between two measures, r_{xx} and r_{yy} are internal consistency scores of two measures and $r_{disattenuated}$ is calculated as follows:

$$r_{disattenuated} = \frac{r_{observed}}{\sqrt{r_{xx} \times r_{yy}}} \quad (1)$$

Participants who were more accurate at pointing at both phases were also more efficient in shortcutting trials, and this relationship is particularly strong in the case of the disattenuated correlations, which correct for internal consistency. However, these results mask individual differences between participants, which are presented in the next section.

Individual differences: Low-spatial participants versus high-spatial participants

A K-means clustering analysis was conducted on three measures (efficiency, Phase I, and Phase II pointing errors) to categorize participants as having low or high-spatial

ability.⁷ Note that two was the optimal number of clusters based on the elbow and the silhouette method (see [Online Supplemental Materials \(OSM\)](#) for additional information). Descriptive statistics and internal consistency for each measure are shown in Table 3, separately for these two groups.⁸

For low-spatial participants, in the desktop study ($N = 37$), the average pointing error before the shortcutting task (Phase I pointing) (86.32° , $SD = 13.01^\circ$), was not significantly different from chance (90°), one-sample t (36) = -1.72, $p = 0.09$, $d = -0.28$, 95% CI = [81.98, 90.66]. Moreover, these participants' pointing performance across trials was not reliable (internal consistency = 0.40). However, their average travel efficiency score was 2.04, which was significantly shorter than the learned route (Efficiency = 2.54), one-sample t (36) = -12.77, $p < 0.001$, $d = -2.1$, 95% CI = [1.96, 2.12], suggesting some ability to take novel paths that were more efficient than the learned route, even though they pointed at chance and their pointing performance was not consistent across trials. Similarly, in the immersive study ($N = 24$), low-spatial participants' pointing performance (85.14° , $SD = 11.16^\circ$) was better than chance, one-sample t (23) = -2.13, $p = 0.04$, $d = -0.44$, 95% CI = [80.42, 89.85], but close to chance. Their pointing performance was also not reliable (internal consistency = 0.12). However, their average travel efficiency (1.89) was significantly more efficient than the learned route (2.19), one-sample t (23) = -5.25, $p < .001$, $d = -1.07$, 95% CI = [1.77, 2.01], suggesting some ability to find shorter paths than the learned route, even though their pointing performance was close to chance and was not consistent across trials.

As shown in Fig. 5, for low-spatial participants, the observed correlations between Pointing Error (Phase I) and shortcutting are not significant (Desktop: $r(35) = 0.00$, $t(35) = 0.02$,

⁶ Data are repeatedly randomly split into two halves 5,000 times. The final internal consistency is the average of the 5000 split-half reliability estimates. (Parsons et al., 2019).

⁷ An alternative way to group participants is using a median split on their pointing performance in the first phase, however, the main conclusions of the paper do not change if we use this method.

⁸ The low-spatial groups were unsuccessful in finding the target on more trials (2.04 trials in the desktop; 2.08 trials in the immersive version) than the high-spatial groups (0.10 trials in the desktop; 0.54 trials in the immersive version). The efficiency score captures these differences as described in the method section, so we only examined the efficiency scores in the following analysis.

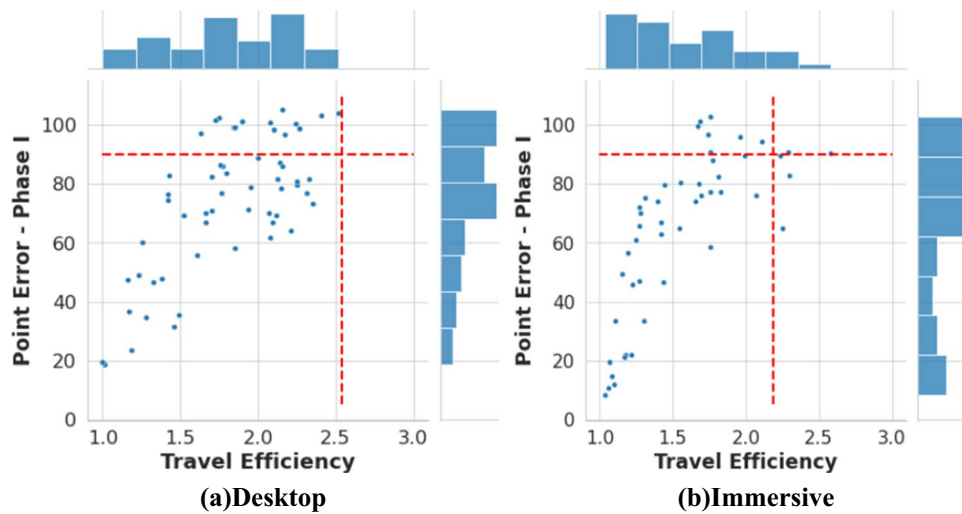


Fig. 4 Scatter plots for the Pointing Error (Phase I) in the direction estimation task Phase I and Travel Efficiency in the shortcutting task in the desktop study (a) and in the immersive study (b). The red horizontal

line indicates the chance level performance: 90°. The red vertical line indicates the average efficiency of taking learned routes on every trial: 2.54 (a: in the desktop study) and 2.19 (b: in the immersive study)

$p = .98$, 95% CI = [-.32, .33]; Immersive: $r(22) = .05$, $t(22) = 0.23$, $p = .82$, 95% CI = [-.36, .44]). These correlations were partially driven by the low internal consistency of both measures, suggesting that individual-level correlation coefficients were attenuated by measurement variance unrelated to true between-individual variances. After correcting for the internal inconsistency of the measure, the disattenuated correlations between the Pointing Error (Phase I) and shortcutting were still not significant (see Fig. 5); that is, low-spatial participants’ pointing performance cannot predict their shortcutting performance.

Table 2 The observed and disattenuated correlation table for all participants

Disattenuated Observed	Travel Efficiency	Pointing Error - Phase I	Pointing Error - Phase II
Desktop			
Travel Efficiency	-	.92***	.97***
Pointing Error - Phase I	.71***	-	.93***
Pointing Error - Phase II	.76***	.78***	-
Immersive			
Travel Efficiency	-	0.86***	0.90***
Pointing Error - Phase I	0.74***	-	0.88***
Pointing Error - Phase II	0.79***	0.78***	-

Values below the diagonal, in the bottom left are the observed correlations and values above the diagonal in the top right are the disattenuated correlations corrected using Eq. (1). *** indicates $p < .001$

For high-spatial participants (Desktop: $N = 20$; Immersive: $N = 24$), pointing performance in the first phase was highly correlated with shortcutting. (Desktop: $r(14) = .60$, $t(14) = 3.03$, $p = .01$, 95% CI = [.15, .85]; Immersive: $r(22) = .75$, $t(22) = 5.35$, $p < .001$, 95% CI = [.50, .89]) with higher correlations after correcting for the internal inconsistency (see Fig. 5). The disattenuated correlations for the high and low-spatial groups were significantly different (Desktop: $Fisher's z = 4.43$, $p < .001$, $Zou's 95\% CI = [-1.28, -0.38]$; Immersive: $z = 7.95$, $p < .001$, $Zou's 95\% CI = [-1.22, -0.44]$).

Note that in the immersive study, the internal consistency for shortcutting was 0.57, which is relatively low. The relatively low internal consistency, in this case, was driven by the close-to-ceiling performance. That is, the variance for each trial was determined by a small number of participants who did not get the perfect efficiency score (efficiency of 1) and so there was limited variance to correlate between trials.

General discussion

We examined the relation between pointing and shortcutting performance after the same egocentric learning experience in two studies, one using desktop VR and the other using immersive VR. The results of these studies are consistent. In both studies, the correlation between shortcutting and pointing depends on participants’ learning ability, as well as the internal consistency and discriminating power of the measures. The high-spatial groups across studies were generally good at both shortcutting and pointing and the correlation between shortcutting and pointing was high for these groups; the low-spatial groups

Table 3 Descriptive statistics and internal consistency for measures

	Spatial Ability	Mean	SD	Min	Max	Skewness	Kurtosis	# of Trials	Internal Consistency
Desktop									
Pointing Error (Phase I)	High	50.36	19.62	18.85	83.04	-0.05	-1.11	27	0.80
	Low	86.32	13.01	61.74	105.26	-0.13	-1.23		0.40
Efficiency	High	1.38	0.23	1.00	1.85	0.20	-0.68	19	0.65
	Low	2.04	0.24	1.52	2.51	-0.24	-0.75		0.01
Pointing Error (Phase II)	High	40.50	14.41	15.89	69.41	0.22	-0.87	27	0.71
	Low	78.82	14.60	52.56	104.67	0.10	-1.10		0.51
Immersive									
Pointing Error (Phase I)	High	44.02	23.09	8.21	75.34	-0.17	-1.46	24	0.85
	Low	85.14	11.16	58.77	102.97	-0.42	-0.24		0.12
Efficiency	High	1.24	0.14	1.04	1.55	0.42	-0.52	24	0.57
	Low	1.89	0.28	1.44	2.58	0.76	-0.22		0.57
Pointing Error (Phase II)	High	28.27	12.04	12.49	51.07	0.41	-1.09	24	0.62
	Low	77.54	14.25	45.35	99.97	-0.70	-0.09		0.47

For the desktop study, only including the corresponding trials 19 trials in the pointing tasks do not change the conclusions. See [OSM](#)

had poor pointing performance but took novel and efficient routes, and shortcutting and pointing were not significantly correlated for these groups.

Relations between shortcutting and pointing were affected by both the discriminability and internal consistency of the measures. In terms of discriminability, we observed a tension between the difficulty of the pointing task for the low-spatial group and the difficulty of the shortcutting task for the high-spatial group (see Fig. 5). The desktop environment was relatively difficult to learn, given the amount and type

of learning experience given in these studies, such that we observed a floor effect for the low-spatial group in the pointing task. The immersive environment was easier to learn, but resulted in a close-to-ceiling effect for the high-spatial group in the shortcutting task. Given the wide range of individual differences in large-scale spatial cognition, we recommend that future researchers examine the discriminating power of their measures and use measures that can distinguish across the full range of ability. They may need to combine multiple measures to assess all levels of environmental learning ability.

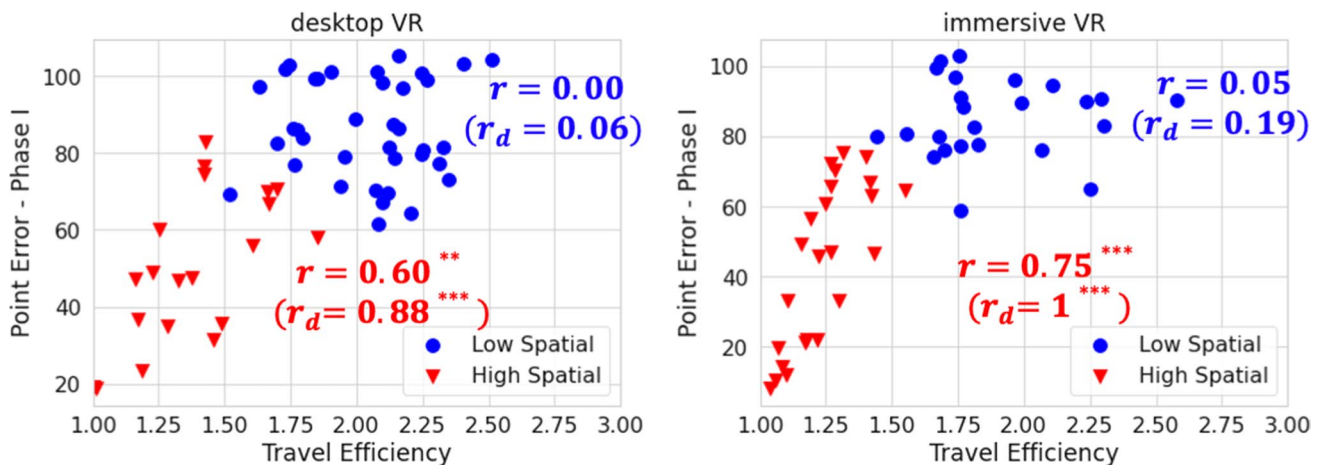


Fig. 5 Scatter plots, observed correlation (r), and disattenuated correlation (r_d) between pointing and shortcutting for high- and low-spatial groups in the desktop study and the immersive study. *Note:* For the high-spatial participants, indicated by the red (triangular) points, the correlations between pointing and shortcutting are almost 1. *** indicates $p < .001$.

However, for low-spatial participants, indicated by the blue (circular) points, the correlations were not significantly different from zero. The low correlations were partially driven by the low internal consistency of the measures. A full correlation table is in the [OSM](#) TS1

Low-spatial participants showed low internal consistency in their pointing and shortcutting performance, while high-spatial participants showed relatively low internal consistency in their shortcutting performance in immersive VR, which attenuated the observed correlation between the two measures (Ackerman & Hambrick, 2020; Hedge et al., 2018; Parsons et al., 2019). The item-level variance may be driven by (1) inconsistent accuracy of mental representations for different locations in the environment (e.g., landmarks near the boundary or aligned with specific orientations may be easier to learn), (2) differential availability of navigational cues in different trials, and (3) participants' differential sensitivity to these cues (e.g., Andersen et al., 2012; Barhorst-Cates et al., 2021; Coutrot et al., 2022; He et al., 2021; Newcombe et al., 2023). Investigating the effect of these factors calls for future studies. Our study highlights that these underlying cognitive processes are masked if researchers do not investigate their instruments by first examining measurement reliability.

These analyses help us advance our understanding of the nature of configural knowledge, specifically on whether this is best characterized as labeled graph knowledge or metrically accurate survey knowledge (Foo et al., 2005; Gallistel, 1990; Kuipers et al., 2003; O'Keefe & Nadel, 1978; Peer et al., 2021; Warren, 2019). Our results show that pointing performance is accurate and is correlated with shortcutting for high-spatial participants, but pointing performance is less accurate and not correlated with shortcutting for low-spatial participants. This suggests that the high-spatial group may have acquired both types of knowledge, whereas the low-spatial group only acquired graph knowledge with this amount of learning experience.

Our pointing task provided only one view of the environment in each trial and did not allow people to look around before estimating the direction. Low-spatial participants' relatively poor performance in pointing might also reflect difficulty orienting themselves in the environment based on this limited information. Future research, using a more immersive pointing measure will help distinguish whether poor pointing performance by this group is due to a poor cognitive map of the environment or an inability to locate themselves in this cognitive map. The present study provides one way of examining the measures, and the key point is that underlying knowledge measured for different people may change if the paradigms and trials are changed.

To conclude, instead of assuming that pointing and shortcutting are interchangeable measures of environmental knowledge, our studies show that it is critical to examine psychometric properties, including reliability and discriminability, before selecting measures or interpreting the correlations. Psychometric properties are largely under-reported in the spatial cognition domain but can advance our understanding of individual differences and should be an important foundation of research on cognitive processes underlying complex spatial tasks.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13423-023-02266-6>.

Acknowledgements This work was supported by the National Science Foundation (NSF-FO award ID 2024633) and the Office of Naval Research (N00014-21-2425). We would like to thank Mengyu Chen and Fredrick (Rongfei) Jin for helping with programming the early demo of the immersive task. We thank Mitch Munns, Bryan Maghen, Tashia Ovais, and Shalmali Patil for helping with the data collection.

Data Availability The datasets generated during and/or analysed during the current study are available in the ConfiguralSpatialKnowledge-Measurement repository [<https://github.com/CarolHeChuanxiuyue/ConfiguralSpatialKnowledgeMeasurement.git>].

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ackerman, P. L., & Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. *Intelligence*, 80, 101440.
- Anastasiou, C., Baumann, O., & Yamamoto, N. (2022). Does path integration contribute to human navigation in large-scale space?. *Psychonomic Bulletin & Review*, 1–21. <https://doi.org/10.3758/s13423-022-02216-8>
- Andersen, N. E., Dahmani, L., Konishi, K., & Bohbot, V. D. (2012). Eye tracking, strategies, and sex differences in virtual navigation. *Neurobiology of Learning and Memory*, 97(1), 81–89.
- Barhorst-Cates, E. M., Meneghetti, C., Zhao, Y., Pazzaglia, F., & Creem-Regehr, S. H. (2021). Effects of home environment structure on navigation preference and performance: A comparison in Veneto, Italy and Utah, USA. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(7), 1401–1414.
- Boone, A. P., Maghen, B., & Hegarty, M. (2019). Instructions matter: Individual differences in navigation strategy and ability. *Memory & Cognition*, 47(7), 1401–1414.
- Caduff, D., & Timpf, S. (2008). On the assessment of landmark salience for human navigation. *Cognitive Processing*, 9(4), 249–267.
- Carpenter, F., Manson, D., Jeffery, K., Burgess, N., & Barry, C. (2015). Grid cells form a global representation of connected environments. *Current Biology*, 25(9), 1176–1182. <https://doi.org/10.1016/j.cub.2015.02.037>
- Chrastil, E. R., & Warren, W. H. (2015). Active and passive spatial learning in human navigation: Acquisition of graph knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(4), 1162–1178. <http://dx.doi.org.proxy.library.ucsb.edu/2048/10.1037/xlm0000082>
- Cohen, J. (1988). The effect size. *Statistical power analysis for the behavioral sciences* (2nd ed., pp. 77–83). Hillsdale, NJ: Erlbaum.
- Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, R. C., Wiener, J. M., Hölscher, C., Hornberger, M., & Spiers, H. J. (2022). Entropy of city street networks linked to future spatial navigation ability. *Nature*, 604(7904), 104–110. <https://doi.org/10.1038/s41586-022-04486-7>

- Cramer, D., & Howitt, D. L. (2005). *The SAGE dictionary of statistics: A practical guide for students in the social sciences* (3rd ed.). London: SAGE.
- Chrastil, E. P., Warren, W. (2013). Active and passive spatial learning in human navigation: Acquisition of survey knowledge. *Journal of Experimental Psychology: Learning Memory and Cognition*, 39(5), 1520–1537. <https://doi.org/10.1037/a0032382>
- Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 195–215. <http://dx.doi.org.proxy.library.ucsb.edu:2048/10.1037/0278-7393.31.2.195>
- Gagnon, K. T., Cashdan, E. A., Stefanucci, J. K., & Creem-Regehr, S. H. (2016). Sex differences in exploration behavior and the relationship to harm avoidance. *Human Nature*, 27(1), 82–97.
- Gagnon, K. T., Thomas, B. J., Munion, A., Creem-Regehr, S. H., Cashdan, E. A., & Stefanucci, J. K. (2018). Not all those who wander are lost: Spatial exploration patterns and their relationship to gender and spatial memory. *Cognition*, 180, 108–117.
- Gallistel, C. R. (1990). *The organization of learning*. The MIT Press.
- 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(5), 877–888.
- He, Q., McNamara, T. P., Bodenheimer, B., & Klippel, A. (2019). Acquisition and transfer of spatial knowledge during wayfinding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45, 1364–1386.
- He, Q., Han, A. T., Churaman, T. A., & Brown, T. I. (2021). The role of working memory capacity in spatial learning depends on spatial information integration difficulty in the environment. *Journal of Experimental Psychology: General*, 150(4), 666.
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50, 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, 34(2), 151–176.
- 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(2), 93–129. <http://dx.doi.org.proxy.library.ucsb.edu:2048/10.1016/j.cogpsych.2005.08.003>
- Kang, S. S., & MacDonald, A. W., III. (2010). Limitations of true score variance to measure discriminating power: Psychometric simulation study. *Journal of Abnormal Psychology*, 119(2), 300–306.
- Kuipers, B., Tecuci, D. G., & Stankiewicz, B. J. (2003). The skeleton in the cognitive map: A computational and empirical exploration. *Environment and Behavior*, 35, 81–106. <https://doi.org/10.1177/0013916502238866>
- Labate, E., Pazzaglia, F., & Hegarty, M. (2014). What working memory subcomponents are needed in the acquisition of survey knowledge? Evidence from direction estimation and shortcut tasks. *Journal of Environmental Psychology*, 37, 73–79.
- Lawton, C. A. (2001). Gender and regional differences in spatial referents used in direction giving. *Sex Roles*, 44(5–6), 321–337.
- Malanchini, M., Rimfeld, K., Shakeshaft, N. G., McMillan, A., Schofield, K. L., ... & Plomin, R. (2020). Evidence for a unitary structure of spatial cognition beyond general intelligence. *npj Science of Learning*, 5, 9. <https://doi.org/10.1038/s41539-020-0067-8>
- McNamara, T. P. (2013). Spatial memory: Properties and organization. In D. Waller, & L. Nadel (Eds.), *Handbook of spatial cognition; handbook of spatial cognition* (pp. 173190, Chapter x, 309 pages). American Psychological Association. <http://dx.doi.org.proxy.library.ucsb.edu:2048/10.1037/13936-010>
- Meilinger, T., Riecke, B. E., & Bühlhoff, H. H. (2014). Local and global reference frames for environmental spaces. *Quarterly Journal of Experimental Psychology*, 67(3), 542–569.
- Muryy, A., & Glennerster, A. (2018). Pointing Errors in Non-metric Virtual Environments. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial cognition XI. Spatial cognition 2018. Lecture notes in computer science*. (Vol. 11034). Springer. https://doi.org/10.1007/978-3-319-96385-3_4
- Newcombe, N. S., Hegarty, M., & Uttal, D. (2023). Building a cognitive science of human variation: Individual differences in spatial navigation. *Topics in Cognitive Science*, 15(1), 6–14. <https://doi.org/10.1111/tops.12626>
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- O’Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Clarendon Press.
- Pagkratidou, M., Galati, A., & Avraamides, M. (2020). Do environmental characteristics predict spatial memory about unfamiliar environments?. *Spatial Cognition & Computation*, 20(1), 1–32. <https://doi.org/10.1080/13875868.2019.1676248>
- Parsons, S., Kruijt, A. W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science*, 2(4), 378–395.
- Peer, M., Brunec, I. K., Newcombe, N. S., & Epstein, R. A. (2021). Structuring knowledge with cognitive maps and cognitive graphs. *Trends in Cognitive Sciences*, 25(1), 37–54.
- Röser, F., Hamburger, K., Krumnack, A., & Knauff, M. (2012). The structural salience of landmarks: Results from an on-line study and a virtual environment experiment. *Journal of Spatial Science*, 57(1), 37–50.
- Ruginiski, I. T., Creem-Regehr, S. H., Stefanucci, J. K., & Cashdan, E. (2019). GPS use negatively affects environmental learning through spatial transformation abilities. *Journal of Environmental Psychology*, 64, 12–20.
- Schinazi, V. R., Nardi, D., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2013). Hippocampal size predicts rapid learning of a cognitive map in humans. *Hippocampus*, 23(6), 515–528.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user’s guide*. Psychology Software Tools, Inc.
- Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In *Advances in child development and behavior*, 10, 9–55. [https://doi.org/10.1016/S0065-2407\(08\)60007-5](https://doi.org/10.1016/S0065-2407(08)60007-5)
- Sorrows, M. E., & Hirtle, S. C. (1999). The nature of landmarks for real and electronic spaces. In *Spatial information theory. Cognitive and computational foundations of geographic information science: International conference COSIT’99 stade, Germany, August 25–29, 1999 Proceedings 4* (pp. 37–50). Springer Berlin Heidelberg.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101. <https://doi.org/10.2307/1412159>
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189–208. <http://dx.doi.org.proxy.library.ucsb.edu:2048/10.1037/h0061626>
- Warren, W. H. (2019). Non-euclidean navigation. *Journal of Experimental Biology*, 222(Suppl 1). <https://doi.org/10.1242/jeb.187971>
- Warren, W. H., Rothman, D. B., Schnapp, B. H., & Ericson, J. D. (2017). Wormholes in virtual space: From cognitive maps to cognitive graphs. *Cognition*, 166, 152–163.
- Weisberg, S. M., & Newcombe, N. S. (2016). How do (some) people make a cognitive map? Routes, places, and working memory.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(5), 768–785.

Weisberg, S. M., & Newcombe, N. S. (2018). Cognitive maps: Some people make them, some people struggle. *Current Directions in Psychological Science*, 27(4), 220–226.

Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2014). Variations in cognitive maps: Understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3), 669–682.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates Publishers.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Practices Statement All experimental materials, raw data and analysis scripts are available on Github (<https://github.com/CarolHeChuanxiuyue/ConfiguralSpatialKnowledgeMeasurement.git>), and none of the studies was preregistered.