The impossible puzzle: No global embedding in environmental space memory

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Marianne Strickrodt
aus Sondershausen, Deutschland

August 2018

Tag der mündlichen Prüfung:	21.12.2018
Dekan der Math.-Nat. Fakultät:	Prof. Dr. W. Rosenstiel
Dekan der Medizinischen Fakultät:	Prof. Dr. I. B. Autenrieth
1. Berichterstatter:	PD Dr. Tobias Meilinger
2. Berichterstatter:	Prof. Dr. Hanspeter A. Mallot
Prüfungskommission:	PD Dr. Tobias Meilinger
	Prof. Dr. Hanspeter A. Mallot
	Prof. Dr. Heinrich H. Bülthoff
	Prof. Dr. Andreas Bartels

Declaration

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel „The impossible puzzle: No global embedding in environmental space memory“ selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

I hereby declare that I have produced the work entitled "The impossible puzzle: No global embedding in environmental space memory", submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, den

Datum / Date

Unterschrift /Signature

Summary

We live in compartmentalized, clustered environments and have to deal with spatial information scattered across rooms, streets, neighborhoods, and cities every day of our life. Yet, we are able to piece this information together in our head, for example, in order to find our way from our flat to our workplace, even when faced with construction work and blocked streets. Furthermore, we can point out the direction to the supermarket to a pedestrian without having direct visual access to it. My thesis is concerned with the question of how our memory for spatial relations of places in navigable space (also called survey knowledge) is actually structured. In four consecutive studies, I contrasted two major theoretical approaches that try to explain how we represent survey knowledge, namely, Euclidean map and enriched graph approaches. Euclidean map approaches assume that spatial locations are represented in a map-like, globally embedded, Euclidean format. Enriched graph approaches propose a partitioned, unit-wise representation of places connected in a network. These local units are not required to be globally consistent. In each study, I used different virtual environments, sometimes single rooms, mostly navigable multi-corridor environments, once even an impossible non-Euclidean environment. Participants learned spatial relations between objects spread across these environments and solved survey tasks afterward (e.g., pointing to object locations from memory). Their performance yielded multiple effects. In short, the most prominent effects were: (1) Pointing latency increased with increasing number of places along the route towards the target, (2) facilitated recall along the direction of the initially experienced path walked within the environment, (3) globally incoherent pointing behavior following the local metrics experienced from place to place, (4) facilitated performance upon alignment with local corridor geometry but also (5) upon alignment with regional geometry and a global main orientation, and (6) decreased pointing latency when pointing beyond regional boundaries. Interpreting these effects jointly implies that human survey knowledge is not represented in the form of a Euclidean mental map embedding all encountered places in a uniform, globally consistent format. Instead, just as the environment we experience, also our memory of it seems to be compartmentalized, consisting of a network of local places connected by directed links that specify how to get from one place to another (rotation and translation) without directly requiring a global calibration. Survey estimates have to be constructed incrementally following this graph structure along the memorized connectivity, thereby relying on the local metrics that enrich the graph entities. These estimates are generally transient but can be retained for

a limited amount of time for aiding subsequent estimates. In addition to the local entities of the enriched graph representation, it seems that general reference directions can be acquired during learning a navigable multi-compartment space. Such a reference direction can be understood as a mental "north", a main direction that is tried to be maintained and propagated across multiple local places and represented supplementary in memory. It might be limited to only a sub-group of local units, thereby forming regional clusters, or it can cover the entire environment that was encountered. Such a general reference direction can aid the coordination of the local memory units during the construction of survey estimates, however, it does not require a global embedding of all place information into a coherent Euclidean map format. In sum, our representation of navigable space seems to be best described as an impossible puzzle where the memorized pieces and connections do not necessarily match up on a global scale.

Acknowledgments

On the windy three-and-a-half-year path through my PhD time I was accompanied by a number of great people whom I would like to thank here. First, I would like to thank my supervisor PD Dr. Tobias Meilinger for his pervasive support and great leadership, his trust in my abilities and his patience, for the freedom he gave me regarding my (not always perfect) time management and for the right amount of pressure at the right time. I will miss our inspiring discussions about experimental design and theory (some of which are even on videotape) and all the fragmented and puzzling idea-snippets in the word comment section. I would also like to thank my second supervisor Dr. Stephan de la Rosa for always taking time for my unannounced questions and discussions, for his diplomatic, fair and open feedback and of course for adopting me and Thomas Hinterecker when Tobias Meilinger left the Institute. I am very grateful that I had the opportunity to work with such amazing supervisors, two very self-conscious, supportive individuals that want nothing but the best for their students. Inspired by the two, I very much hope that in the future I will be able to become a trustworthy, smart and strong supervisor and leader myself.

I am very grateful that I had the opportunity to be part of the Social and Spatial Cognition group. Since some of the group members finished their PhDs and left the institute already a few months ago I am already now missing Aurelie Saulton's spirit, Laura Fademrecht's laughter, and Dong Seong-Chang's exceptional character. Especially, I want to thank my PhD colleague Thomas Hinterecker, who is so smart and disciplined that he handed in his thesis three days before me even though he started his PhD a few months after me—this mental pressure made me run faster (in the positive sense). Thanks to him I don't have to grapple with SPSS anymore but can now curse R. I very much appreciate having a friend around who is so supportive, calm and relaxed and so honest that he does not hesitate to tell you that your Unity skills have not yet reached level four of five.

I also want to express my gratitude towards Prof. Dr. Heinrich Bülthoff for offering me this PhD position at the Max-Planck-Institute for Biological Cybernetics in Tübingen, for supporting each and every one of his researchers, students and staff, and especially, for having created such a thriving, familiar working climate and an atmosphere of psychological security.

Special thanks go to Joachim Tesch who basically made my experiments in the tracking hall run smoothly. As one of the calmest and patient people I've met in my life, never grown weary of explaining to me how to recalibrate a head-mounted display, he was able to prevent me from pushing the wrong buttons and from bursting the equipment into flames. Also, without his remote support during my internship, I would have never been able to get the tracking system in Providence up and running.

I would also like to acknowledge Prof. Dr. William H. Warren for his contribution to my impossible maze study with numerous skype meetings and discussions during my visit to his lab. This collaboration was and still is a great experience. Furthermore, I would like to thank Prof. Dr. Hanspeter A. Mallot and Prof. Dr. Andreas Bartels, who were part of my advisory board, for steering this work in the right direction with their critical comments and fruitful discussions.

Alongside all the support and input from the scientific side of my life I, of course, had a lot of backing from my family and friends mostly from to long a distance between us. I very much want to thank my family for their faith in me, for always having an open ear for me, for always being interested to know "how the school is going" (Yes, school is over now). I very much appreciate their creativity when trying to "sell" and explain my research topic to friends and colleagues, for example, by stating that I am working on something related to the spatial arrangement of grocery items in the supermarket (I think this helps to grasp the idea). I am also tremendously grateful for the mental support of my friends from home (too far away), Ramona, Katharina, Anne, and Luise, who did never give up on me even in high-workload phases with close to zero sign of life from my side. Being with you always charges my batteries from all the laughing and good spirits. Especially, I want to thank Markus Klein for being on my side, for his support and his patience particularly in high workload-phases, for letting go when I left to Providence and for his love during the PhD time and beyond.

Table of contents

1 Introduction

The environment we are living in prompts us to navigate through many local entities. Consider your everyday life activities, for example, getting out of bed, getting ready in the bathroom, eating breakfast in the kitchen, then leaving the house to get to work. Depending on the distance to your workplace you might walk, cycle, drive by car, bus or train, but in essence, you are traveling through a number of interconnected local environments, such as streets, alleys or open places. At your workplace again, you are confronted with a large number of rooms, corridors or halls often spread across multiple floors that are making up the workplace building. On your way back home in the afternoon you might take different routes, stopping by at the grocery shop or at a friend's house. Due to the nature of this navigation you will need to relate spatial information that could not be explored all at once but instead were experienced distant in space and time. Albeit living and operating in such a fragmented world, humans are able to find their ways (normally) without getting lost and grasp spatial relations between the visited places. We know which turns to take to reach the bakery shop in our neighborhood. When standing in front of the fridge in our kitchen we can point directly (straight-line direction) to the toilet in the bathroom although not in sight. In case of construction work in a frequently traveled street we will be able to find a novel route to our goal through streets never traveled before. Thus, there must be a mental conceptualization of the space we navigate. The question of how we mentally represent the navigable, compartmentalized space we live in is addressed in the spatial cognition research already for many years. The aim of the present doctoral thesis was to add a few more buildings blocks to the existent rich literature and to broaden our understanding of how we make sense of all the locally confined places we encounter. Precisely, I sought to examine the structure of navigable space memory, to test whether the compartmentalized space we experienced is likewise stored in a compartmentalized, memory-unit-by-memory-unit fashion, and—if this is the case—to understand how relations between these mental units are characterized in our memory structure.

1.1 Spatial memory of local environments

To approach concepts of how humans represent large, navigable space I would like to start with a simpler case. Imagine a locally confined space where you can see everything that surrounds you, a space clearly circumscribed by opaque walls, for example, your office at work. Most likely it contains a desk, a chair, a bin, books, and folders arranged on shelves, a computer screen and a drawer (maybe multiple of these items in case you share your office) arranged within the room. How do we make sense of such a spatial surrounding in order to represent it mentally? In the natural case we understand a space and keep track of it relative to our moving body mainly based on visual and idiothetic information (e.g., Rieser, Pick, Ashmead, & Garing, 1995). Visual cues that underly our 3D percept of the environment are binocular disparity (seeing the environment from two slightly distant vantage points, the right and the left eye) (e.g., Van Den Berg & Brenner, 1994), surface texture (distortions like convergence of presumably parallel lines if textured surfaces are viewed from perspective) (e.g., Saunders & Backus, 2006), optic flow (movement of objects across retina, e.g., expansion of a dot field indicating forward movement) (e.g., Lappe & Rauschecker, 1995), motion parallax and occlusion (change of view upon an object during movement). Such 3D depth cues help us to understand the layout of objects (including distance to and relations between objects), the geometry of the environment and our relative position within.

While many of the visual cues can help understand the space around us without ever walking a single step (i.e., static cues), active or passive movement first enhances the visual cues (as it introduces stronger visual change in the scenery) and additionally provides idiothetic information of our body moving through space (e.g., Mittelstaedt Horst, n.d.; Mittelstaedt & Mittelstaedt, 2001). Being passively pushed around in a wheelchair (blindfolded vs. seeing)—as done for example in studies investigating the interplay of visual and idiothetic information for learning space (e.g., Chrastil & Warren, 2012; Waller & Greenauer, 2007)—allows for the use of vestibular afferents from the inner ear otoliths that sense translational and rotational acceleration (i.e., inertial idiothetic cues) (e.g., Israel & Berthoz, 1989). Actively walking additionally involves podokinesthetic (or podokinetic) signals (e.g., Weber, Fletcher, Gordon, Jones, & Block, 1998), which include, on the one hand, proprioceptive afferents about the position of body parts such as the feet, the legs and the hip and their displacement relative to the floor, and on the other hand efferent motor plans that determine the path and efference copies (i.e., substratal idiothetic cues).

The tight coupling of both visual and idiothetic cues enable us to grasp the spatial arrangements of objects within our immediate surrounding and to identify our position within. The more cues available, the better the metric understanding of space and one's own movements within (e.g., Klatzky, Loomis, Beall, Chance, & Golledge, 1998). Nevertheless, even when walking a meter through your office and rotating your body a bit with eyes closed you will be able to keep track of the position of your chair, your screen, and your desk, or even the corners of your office relative to your changing location. The process of keeping track of locations within ones surrounding (whether in sight or not) during self-motion is referred to as spatial updating (e.g., Holmes & Sholl, 2005; Wang & Spelke, 2000). Involved in this is the process of keeping track of one's self-motion over time relative to a fixed starting position, also called path integration (e.g., Loomis et al., 1993). Updating is considered to be highly automatic (e.g., when moving blindfolded in a room it is harder to recall the original object positions at the start compared to recalling the factual, nonvisible target location, which changed during blind motion; Farrell & Robertson, 2000; Martin & Thomson, 1998), but also limited in capacity (e.g., drop in performance when trying to update the position of more than six objects blindfolded; Hodgson & Waller, 2007; see also Wang et al., 2006).

The updating of object locations during locomotion is considered to be part of a highly precise but transient online spatial system by many theories (e.g., Easton & Sholl, 1995; Gallistel, 1990; Huttenlocher, Hedges, & Duncan, 1991; McNamara, 2003; Mou, McNamara, Valiquette, & Rump, 2004; Wang & Spelke, 2000), and contrasts a second system, namely, a more permanent but also much coarser offline system based on long-term memory representations of space (e.g., Mou et al., 2004). Support for this two-system approach comes for example from a study by Waller and Hodgson (2006) that showed how direction estimates to objects within a room are differently affected by blindfolded disorientation depending on whether participants had to perform an egocentric pointing task (i.e., "Where is object X relative to your current location?") or a judgement of relative direction task (also called JRD, i.e., "Imagine you are at position X, facing object Y, please point to object Z."). While egocentric pointing performance decreased with increasing disorientation, JRD tasks' performance increased with increasing disorientation. The drop in performance for the egocentric pointing task is interpreted as losing track of the updated object locations within ones surrounding that was supported by the transient online system. The increase in performance in the JRD task instead can be explained by reverting to the long-term representation of the object layout in this task (i.e., the offline system), which initially interferes with the current information of one's orientation within the environment, but decreasingly so the more a person is losing track of one's orientation in space. Furthermore, performance was more precise in the egocentric task (online system) compared to the JRD task (offline system). Such opposing performance patterns support the dissociation of the online and the offline system.

A great number of studies investigating the enduring long-term representation found that this memory is orientation dependent (e.g., Kelly & McNamara, 2008; Meilinger & Bülthoff, 2013; Mou & McNamara, 2002; Shelton & McNamara, 2001). These studies often used object arrangements learned in a single room and had participants perform a judgment of relative direction (JRD) task afterward. For example, in one trial participants were instructed to “Imagine you are at position X, facing object Y, please point to object Z.”, while in another trial the to-be-imagined orientation changed by instructing to face another object: “Imagine you are at position X, facing object W, please point to object Z.”. Like this over many trials a whole range of to-be-imagined body orientations was queried. Indeed, pointing was found to be faster and more accurate when participants were aligned with certain orientations (of which one orientation often elicited the best performance) compared to being oriented otherwise. The often replicated sawtooth-pattern of performance over the tested orientations is typically explained by encoding object layouts relative to one or more orthogonal reference axes of a common reference system, as if object locations are assigned unique coordinates in a mental cartesian coordinate system (McNamara, Sluzenski, & Rump, 2008; Mou et al., 2004; Shelton & McNamara, 2001)[1]. Being aligned with these axes allows for an effortless retrieval of coded information. Drops in performance for oblique orientations are thought to reflect costs of mental transformations as one must mentally align the oriented representation stored in long-term memory with one’s current orientation (e.g., McNamara et al., 2008; Meilinger, Berthoz, & Wiener, 2011). Albeit transformational costs resulting from the alignment process, all inter-object relations should be directly represented allowing for a simple read-out of coordinate values and computation of difference vectors between objects. According to Shelton and McNamara (2001), an environment is scanned to identify a conceptual “north”, a dominant reference direction to anchor the spatial mental reference system to. Multiple factors have been found to be used as a reference direction thereby setting the mental reference frame orientation, for example, egocentric experiences (e.g., first perspective taken within the environment) (e.g., Kelly & McNamara, 2008; Rieser, 1989), salient layout intrinsic cues (e.g., multiple objects forming rows, columns, orthogonal and symmetric arrangements) (e.g., Kelly & McNamara, 2008; Richard & Waller, 2013), salient layout extrinsic cues (e.g., rectangular room forming an

[1] Results from Street and Wang (2014) suggest that only one reference axis is represented in memory. Enhanced performance in contra- and orthogonally aligned orientations relative to this reference orientation seem to mirror lower transformation costs compared to obliquely aligned orientations, instead of a representation of additional reference axes.

elongated geometry) (e.g., Shelton & McNamara, 2001; Valiquette & McNamara, 2007) and even instructions (e.g., instructed to pay attention to a specific perspective) (e.g., Mou & McNamara, 2002). Such reference frames store object-to-object relations with respects to environmental elements and independent of the navigator's body and are thus per definition allocentric reference frames, which are conceptually different from egocentric reference frames (e.g., Klatzky, 1998). Egocentric reference frame code self-to-object relations, thus, the relative location of objects with respect to the observer.

1.2 Memory for vista vs. environmental space

Coming back to the example of your workplace now imagine leaving your office to join a meeting in another room—what happens to the items within your office and all the new items you encounter on your way to the meeting room (e.g., a printer, cabinets, plants, armchair, and couch)? In the literature there have been many attempts to conceptualize the space we encounter, often based on dimensions like size, or locomotion (for a comprehensive summary see for example Freundschuh & Egenhofer, 1997). One of the taxonomies often reverted to in the spatial cognition literature is the one by Montello (1993). He differentiates figural, vista, environmental and geographical space. Figural space is generally visible at a glance. It can be both small 3D object spaces that are reachable and manipulable as well as 2D pictorial representations like pictures or downscaled representation of larger spaces such as maps. Geographic space is defined by its size, which is too large to be experienced by locomotion but instead must be learned via a symbolic, downscaled representation of the geographic scale (e.g., a map; this symbolic representation ultimately renders this geographic space a figural space). Environmental spaces are defined as spaces one must navigate through in order to experience it (but obviously smaller in size compared to geographical space). This includes, for example, buildings (e.g., your workplace building), a neighborhood, or a city. Environmental spaces contrast thereby nicely the concept of a vista space, a space that can be learned from a single vantage point such as a room (e.g., your office), a corridor, a single street, a town square or alley.

Knowledge of environmental space was categorized by Piaget and Inhelder (1969) into different formats, namely landmark, route and survey knowledge (see also Siegel & White, 1975). Landmark knowledge was described as memory for salient reference points that allow to identify specific locations (e.g., a church, a river, a unique feature in the environment) and is typically tested with recognition tasks (e.g., Caduff & Timpf, 2008; Presson &

Montello, 1988; Siegel & White, 1975). Route knowledge refers to memory about correct actions one must take at decision points (i.e., intersections) in order to pursue the correct path towards a goal. The relative position of the goal itself is not represented in route knowledge. Instead route knowledge is often referred to as simple representation of interlinked stimulus-response associations: A cue/landmark at an intersection is associated with a motor response such as "turn right" and followed by another cue/landmark which again is associated with another motor response (e.g., O'Keefe & Nadel, 1978; Siegel & White, 1975; M. Strickrodt, O'Malley, & Wiener, 2015; Wolbers & Wiener, 2014). Route knowledge is typically measured by wayfinding tasks, for example, confronting participants with one or more decision points asking for the correct direction to take to follow a previously learned path (e.g., Janzen, 2006; Marianne Strickrodt, O'Malley, & Wiener, 2015), or by simply querying the sequence of motor responses that must be taken along a route from start to goal (e.g., Meilinger, Frankenstein, & Bülthoff, 2013). Finally, survey knowledge refers to the memory of the relative location of places/objects within the environment irrespective of the path one must take to get there. This includes knowledge about distances and relative directions, for example the knowledge that, when standing at your entrance door at home, the bakery shop is located roughly 45° towards the left relative to your current view and in 600m distance when considering an as-the-crow-flies straight line, whereas the next bus-stop is located roughly 400m right behind you in a parallel, currently non-visible street. Survey knowledge is the form of spatial memory that enables us (in contrast to for example route knowledge) to take novel shortcuts to a known location across previously untraveled terrain, to point in a straight line towards a distant goal currently not in view or estimate its distance and it is measured with the appropriate tasks. In the following sections I will outline different theoretical concepts that try to explain how survey memory is represented. The question I pursue in my thesis is which of these concepts actually captures the underlying memory structure that guides our survey behavior.

1.3 Theories and concepts of survey knowledge

1.3.1 Euclidean cognitive maps

For a long time already cognitive scientists hold a debate about how survey knowledge is structured. Already 1948 Edward C. Tolman introduced the term *cognitive map* to describe a comprehensive long-term representation of environmental relationships formed by rats that he found to be able to identify novel shortcuts, a straight-line direction from a start towards a target place that was previously traveled to non-directly via multiple corridors. The concept Tolman described back then constituted a contradiction to the conviction at that time that rats merely act in space based on simple stimulus-response strategy, thus, they only replicated movements previously executed when learning the path the first time. In the time since the notion of a *cognitive map* proved to be very influential. It was often referred to and refined in the human spatial cognition literature and often utilized and understood in its literal sense: Spatial relations of objects and places within our environment were often suggested to be represented in the structured format of a 2-dimensional map (e.g., Evans, 1980; Gallistel, 1990). According to Thorndyke and Hayes-Roth (1982) places are memorized relative to a fixed coordinate system where also Euclidean straight-line distances between objects are represented. Poucet (1993) uses the following wording: “As a matter of fact, the buildup of a spatial representation implies that information sequentially acquired (as a result of an animal's movements) be integrated into a maplike structure allowing for simultaneous access to all relevant information” (p.163). Likewise, Ishikawa and Montello (2006) explicate the following: “For survey maps to emerge, routes need to be metrically scaled and interrelated into a global allocentric reference system. In other words, places and routes learned during separate travel experiences are integrated and interrelated with each other in a common frame of reference” (see also Montello, 1998; Siegel & White, 1975). Thus, these concepts of survey knowledge as a cognitive map are in strong accordance with the model of reference frames suggested by McNamara and colleagues (e.g., McNamara et al., 2008; Mou et al., 2004; Shelton & McNamara, 2001). However, whereas the model of reference frames is mostly used to explain pointing patterns in vista spaces, a similar representational structure is assumed for larger, navigable environmental spaces by cognitive map approaches. Nadel (2013) emphasizes “Maps are enormously powerful informational tools because every point on the map is related to every other point within the reference frame of the map” (p. 166). A process that might allow populating a metric map was proposed by Gallistel (1990; Gallistel & Cramer, 1996). When moving through space the path integration system registers the bodily displacement relative to a cardinal y-axis (a mental “North”) and its orthogonal counterpart, the x-axis. For salient places or landmarks that are encountered their allocentric coordinates relative to the starting location (x,y) = (0,0) are computed and entered in the metric map. Iteration of this process allows to fill up the globally consistent coordinate system. Figure 1 illustrates how the external world (left) might be represented in a Euclidean coordinate system (middle). Common to all these cognitive map approaches is that humans possess a metric map of their environment, which represents spatial locations relative to a reference system with a designated main orientation. Further, such representations need to conform with the axioms of Euclidean geometry, which will be further expounded in a bit[2].

It is argued and supported by numerous studies that rodent and mammal brains possess the ability for such a global metric embedding. In the past years we came to unravel and understand the neural architecture and functioning of spatial memory based on numerous neuroscientific experiments, ranging from single cell recordings in rats, lesion studies in humans, cell recordings from epilepsy patients, through to fMRI studies that observe BOLD-signal patterns that correspond to the more direct recordings of cells. Although there is no topological mapping between places in the real world and neuronal ensembles, different cell types have been discovered that show activity patterns matching definable spatial properties and were therefore interpreted as the neural correlates of an inertial coordinate system that support the encoding of metric maps from path integration (e.g., Derdikman & Moser, 2011; McNaughton, Battaglia, Jensen, Moser, & Moser, 2006; O'Keefe & Nadel, 1978). Place cells are typically found in the hippocampus and appear to represent individual places in space (e.g., O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978). Having a mouse explore a laboratory box while recording a collection of hippocampal place cells reveals that single cells are sensitive to specific places within that box. Independent of the view-point every time the mouse revisits this place the same place cell becomes activated while staying silent in other parts of the box, thus, representing the animal's allocentric location within the environment. In contrast, grid cells are firing repeatedly across the entire box in a regular fashion, forming a six-fold, hexagonal activation pattern. Associated with this regularity, information about distance and direction traveled can be provided. Therefore, grid cells located in the entorhinal cortex are considered to code self-motion-based processes aiding path integration (e.g., McNaughton et al., 2006; Moser, Kropff, & Moser, 2008). Information about the navigators' direction is carried by head direction cells. Irrespective of the navigators' position in space the firing rate corresponds to the viewing direction (Taube et al., 1990; Taube, 2007). In humans, evidence for place cells in hippocampus and

---

[2] A slightly different cognitive map approach by Sholl (2001 Easton & Sholl, 1995; Sholl & Nolin 1997) suggests that spatial relations of environmental space are stored in a cognitive vector space specifying object-to-object relations on the local place-to-place basis. The model holds that the object-to-object relations are not embedded in a cartesian system with a predefined "north" axis but represented in an orientation-independent fashion. Albeit not explicitly stated Sholls model seem to assume that this vector space is metrically embedded in a globally coherent format. However, unfortunately she owes and explication.

grid cells in the entorhinal cortex comes from cell recordings in neurosurgical patients (Ekstrom et al., 2003; Jacobs et al., 2013; Jacobs, Kahana, Ekstrom, Mollison, & Fried, 2010) and less directly also from interpreting BOLD-signal patterns in fMRI studies (e.g., Doeller, Barry, & Burgess, 2010; Horner, Bisby, Zotow, Bush, & Burgess, 2016; Kunz et al., 2015). Further, fMRI studies showed corresponding patterns for the sensitivity for single places and facing direction (similar to the function of place and head direction cells) in other medial brain structures (retrosplenial complex and the presubiculum) (e.g., Marchette, Vass, Ryan, & Epstein, 2014; Vass & Epstein, 2013). Using these cell types in concert is thought to be the basis for assigning coordinates to salient places visited based on the visual and idiothetic information of translations and rotations through space (e.g., Gallistel & Cramer, 1996; McNaughton et al., 2006; O'Keefe & Nadel, 1978).

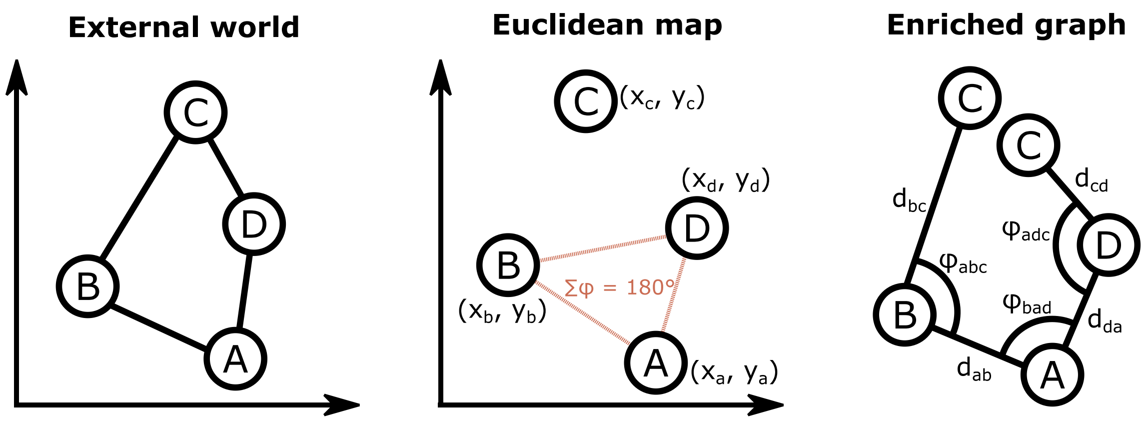


Figure 1. **Left:** External world layout of four places in an environmental space that are learned successively from A to B, C, D and then back to A. **Middle:** Euclidean map representation. Each place is assigned to unique coordinates in a mental coordinate system. Like this angular and distance information can be read out immediately. This representation can be subject to distortions and deviate from the external world. However, it must be globally embedded obeying Euclidean metric postulates. **Right:** Enriched graph representation. Angular and distance information is stored on the local place-to-place level without being globally embedded. Survey estimates are based on the successive recall of these local metrics, which might lead—without conflict—to different estimates (see C) depending on the mental path along the graph. Modified from study 4.

In sum, there are several assumptions underlying very strong and rigid conceptions of cognitive maps (see also Figure 1 middle): The representation should be allocentric (i.e., independent of the navigators' position in space) and places should be stored relative to a common coordinate system while preserving properties of the world in a metric format, thereby upholding Euclidean axioms. This in turn should enable an all-at-once readout of the spatial information stored within the cognitive map since all information should be readily accessible in a common representational format. Metric postulates a cognitive map must uphold are symmetry, positivity, and triangle inequality (e.g., Beals, Krantz, & Tversky, 1968; McNamara & Diwadkar, 1997; Warren, Rothman, Schnapp, & Ericson, 2017). Symmetry is met when survey estimates between pairs of places are reversely commutative, meaning that the distance estimated from A to B must be the same as the estimate from B to A. Positivity is achieved when the distance between any point and itself is zero (i.e., there is only one clearly defined location in the map a place of the external world is assigned to) and when the distance between any two points is larger than zero (i.e., distinct places cannot overlap in the cognitive map). Triangle inequality refers to the relationship between any three points (e.g., A, B, C) on the map. More precisely, adding up the distance between points A and B and between points B and C must always be larger or equal to the distance between points A and C. Additionally, an assumption specific to Euclidean space is that the inner angles of such a triangle should sum up to 180° (see red lines in Figure 1 middle). In sum, this means that the translational and rotational metrics of a Euclidean cognitive map might render a distorted version of the external world but must comply with the limits of the metric postulates. In the following the rigid interpretation of a cognitive map being a full metric map will be referred to as the *Euclidean map* approach.

1.3.2 Non-hierarchical enriched graph representations

Contrasting such all-encompassing, metric cognitive maps alternative approaches have been proposed that combine topological knowledge of connectivity between places and local metric information in a non-hierarchical graph or network structure thereby eliminating the need for a global metric embedding (e.g., Chrastil & Warren, 2014; Meilinger, 2008; Warren et al., 2017). Such graph representations are conceptualized in the form of nodes, which represent places within the environment, and edges, which define links between these places. To enable survey estimates metric information is stored in these graph representations but on the local level of single nodes and/or edges. In the network of reference frame proposed by Meilinger (2008) edges interlinking nodes define the translational and rotational perspective shifts that are necessary to get from one node/place to another. Each node is represented in the form of a reference frame encompassing spatial information from the local environment such as a corridor or room. The labelled graph model (e.g., Chrastil & Warren, 2014; Warren et al., 2017) deviates slightly in that nodes are not specified as local reference frames and further that edges are labelled with translational information (i.e., distance) and nodes are labelled with rotational information (i.e., junction angles) between outbound edges to two or more other nodes. Estimating survey relations between distant places in both models can be achieved by simple vector addition from one's current position along the graph towards the target. The network of reference frames explicates this retrieval process in detail by stating that a navigator imagines the upcoming places along the graph to the target successively, thereby forming a transient, common reference frame that encompasses current and target location. This process is referred to as building a *mental model*. Despite these differences, the important commonalities are that path integration enables recording of walked paths and turned angles in space and that this information is stored in a piecewise fashion as local metrics that define how to get from one node to the adjacent node(s). This piecewise representation can then be used to point to or plan a short-cut to a distant target by successive activation and processing of all the local information that resides between one's current location and the target online during the retrieval process within working memory. Refer to Figure 1 right for a visualization of an enriched graph representation. These models maintain that each local metric can be independently distorted and therefore spatial knowledge must not be geometrically consistent on a global scale, meaning that Euclidean axioms of symmetry, positivity and triangle inequality must not be upheld. This implies, for example, that in a circular environment as the one illustrated in Figure 1, which consists of the interlinked places A, B, C and D and from D another link/edge to A, survey estimates from A to C can vary according to the local metrics that are processed for making the estimate. The estimate can be done by processing and incorporating the translational and rotational information stored via the graph sequence A-B-C or via the graph sequence A-D-C. As the deviations from the real external space are independent for every local metric a person can end up pointing to different locations albeit queried the same target. Throughout my thesis I will use the term *enriched graph* models to refer to the globally non-embedded non-hierarchical graph representations just described.

So, while the Euclidean map approaches assume that spatial information gathered across multiple interlinked rooms or corridors (i.e., vista spaces) are brought into a coherent geometrical format by imposing a single cognitive map covering all these local entities, the enriched graph representations treat local places as discrete entities which are connected to one another without the need for forming a coherent global map-like picture of the environment. Indeed, there are also attempts to unite both models. For example, the space-graph approach by Mallot and Basten (2009) specifies possible variations of the level of metric embedding of a place graph layer. They define three possible types of graphs containing metric information that cover both the Euclidean map and the enriched graph models described here so far. A graph representation enriched by local metrics stores globally inconsistent local metrics (hence, an *enriched graph* in our terms), which can—but don't have to—be optimized with the help of triangulation to conform with global metric postulates. This metric embedding is done either by correcting local inconsistencies (which is still coordinate-free; compare to Easton & Sholl, 1995; Sholl, 2001; Sholl & Nolin, 1997) or by assigning optimized values within a coordinate system (hence, becoming a *Euclidean map* in our terms). The most complete case in the space-graph approach is the full metric map where each point in the coordinate system is assigned an identity (i.e., is this coordinate occupied by an object/place or not). The space-graph further proposes the flexible use of various labels (e.g., local translation and rotation information, views, global metric coordinates, functionality labels) to enrich the spatial knowledge of the graph. Assuming such a flexible model allows accounting for many different findings in the spatial cognition literature. In my thesis, however, I question whether we actually need to assume that global metric embedding occurs and that Euclidean maps are stored.

1.3.3 Hierarchical representations

Besides non-hierarchical, single layer representations in the form of enriched graphs or Euclidean maps, alternative approaches postulate that multiple layers can be formed to represent environmental space. These models often extend and/or combine the non-hierarchical models by postulating hierarchies (e.g., Mallot & Basten, 2009; McNamara et al., 2008; Meilinger & Vosgerau, 2010; Stevens & Coupe, 1978; Wiener & Mallot, 2003). Wiener and Mallot (2003) for example proposed a hierarchical graph. In addition to the local level where each node represents a visited place (fine level), a second super-ordinate graph level is added where nodes represent regions comprising multiple places (coarse level). The coarse level is consulted when targets are located in another region relative to one's current position. A similar idea to the model of Wiener and Mallot (2003) has been stated already by Stevens and Coupe (1978) who proposed a hierarchical semantic network containing concepts (i.e., place and region nodes) connected by arrows enriched by spatial relationship labels (e.g., "east from", "1000m"). These theories share the assumption that spatial entities are subsumed to form a new super-ordinate entity. The format of these super-ordinate entities of regions or clusters has been proposed to be rather non-spatial in the form of semantic or conceptual labels (e.g., Hirtle & Jonides, 1985; Hirtle & Mascolo, 1986), or spatial, covering aspects from topological understanding of connectivity and containment (e.g., Stevens & Coupe, 1978; Wang & Brockmole, 2003b, 2003a; Wiener & Mallot, 2003) via the computation of an overall location-independent reference direction attached to each local place (e.g., Poucet, 1993) through to metrically embedded representations such as an additional global reference frame integrating multiple local spaces and defining a global reference direction (e.g., Greenauer & Waller, 2010; McNamara et al., 2008). The latter assumptions, integrating multiple vista space information under a single, superordinate reference frame indeed allows for the proposal of a hybrid model encompassing both the enriched graph and Euclidean map model at the same time. While local vista spaces are still treated as individual memory units, which for example possess a local reference frame where all objects within that vista space are embedded, a metric embedding of spatial information from multiple vista spaces into a Euclidean map can be assumed in addition on a higher level of hierarchy.

1.4 Linking theory to observation

Having portrayed different concepts of how environmental space might be structured in memory, in the following I now want to summarize experimental findings that support or violate the discussed models. So, what do we know about human survey knowledge? What happens when leaving a room and how do we perform survey tasks such as point to distant targets? We are generally well able to build up survey knowledge in environmental spaces. Maps can be helpful during this process but not necessary, instead survey tasks can be solved based on direct egocentric navigation experience alone (e.g., Meilinger, Frankenstein, Watanabe, Bülthoff, & Hölscher, 2015; Richardson, Montello, & Hegarty, 1999; Rossano, West, Robertson, Wayne, & Chase, 1999; Taylor, Naylor, & Chechile, 1999). Even when learning two routes in two neighborhoods separately, between-route pointing was usually possible with sufficient accuracy instantly after participants were exposed to the connecting route between the neighborhoods (e.g., Han & Becker, 2014; Schinazi, Nardi, Newcombe, Shipley, & Epstein, 2013; Weisberg, Schinazi, Newcombe, Shipley, & Epstein, 2014; but see Ishikawa & Montello, 2006). However, when taking a closer look at the performance in survey tasks it becomes apparent that survey knowledge of environmental space seems to be subject to strong biases and clustering effects.

1.4.1 Evidence for local nodes of enriched graphs

It was found that updating of object locations seems to be impaired when leaving a room (Wang & Brockmole, 2003a) as if participants suddenly lost track of what was previously updated when separated by a barrier. Thus, automatic self-to-object updating seems to be concentrated on the immediate, local environment and ignores targets lying behind a visible border (Avraamides & Kelly, 2010; Kelly, Avraamides, & Loomis, 2007; Wang & Brockmole, 2003b).

Also, long-term representations seem to be affected by opaque boarders typically accompanied by constraint visibility of objects within the environment. From the animal literature we know that place cells, albeit typically identifying a single unique place, seem to be re-used and assigned to geometrically novel locations when mice are leaving one compartment (i.e., local environment) and enter an adjacent compartment (i.e., another local environment) (e.g., Derdikman et al., 2009; Fenton et al., 2008; Spiers, Hayman, Jovalekic, Marozzi, & Jeffery, 2015). Further, grid cells have been observed to remap each time a mouse enters a new corridor, thereby disrupting the regular activity pattern of a six-fold symmetry normally exhibited within local environments (e.g., Derdikman et al., 2009; Fyhn, Hafting, Treves, Moser, & Moser, 2007). In a study by Meilinger, Riecke and Bülthoff (2013) participants had to learn object locations in a complex, multi-corridor, virtual environment and later had to solve a straight-line pointing task from various locations within that environment to the other objects encountered during learning. Similar to the method applied in the JRD tasks, participants were bodily aligned with one of eight possible orientations each time they were teleported to a new location within the virtual world. Pointings to the target had to be made relative to their current location and orientation. Note that JRD tasks are typically done purely mentally without visual input, whereas virtual setups like the one used by Meilinger and colleagues (2013) allow to teleport participants to any point in virtual space and to provide the corresponding visual scenery. Participants performance indeed depended on the predefined body orientation during pointing. Importantly, best pointing performance was achieved when participants were aligned with the first view they experienced within every single corridor, so when looking along the elongated corridor walls, compared to for example looking against the wall or being oriented obliquely to the corridor geometry. These results suggest that local reference frames have been formed by the participants, one reference frame for each individual corridor and access of spatial information beyond that corridor was still relying on this local reference frame (see also Werner & Schmidt, 1999, for a slightly different approach unfortunately confounding local and global reference directions but suggesting similar conclusions). This contrasts findings from vista space studies involving learning of a single layout. Typically, a single main orientation is found which upon alignment facilitates judgment of relative direction to other objects, and this is found independent of where a participant has to judge spatial relations from (e.g., Shelton & McNamara, 2001). These findings are usually interpreted as the formation of a single reference system framing the relative location of all objects learned.

Taken together such results indicate that visual separation by walls affects the processing of spatial locations across locally confined environments. The formation of local, distinct reference frames, one for each corridor learned, further suggests that vista spaces are treated like individual memory units. Consistently, Marchette, Ryan and Epstein (2017) showed that, albeit having a relative exact memory of the location of an object within a room, confusion errors occur when participants must select among similar rooms potentially containing the target object. Distinct memory units for visited vista spaces can be well explained by enriched graph representations that emphasize the importance of local environments as molds for the formation of individual memory units (i.e., one node in the graph for each place visited). In contrast, a Euclidean map representation should store all locations, even if experienced across multiple vista spaces, within one coherent metric format. Realizing such a format involves the need to keep track of non-visible targets during learning and the formation of a single global reference frame with a coherent main orientation stretching across the entire environmental space (except for Sholl, 2001, who assumes orientation independence). The results summarized in this subchapter are not sufficient to conclude that global metric embedding into a Euclidean map does not occur. For example, assuming a hierarchical representation where local units are stored but globally embedded into a Euclidean map on a superordinate level is still a valid proposition to account for these results. Nevertheless, the summarized findings are a first indicator that leaving a vista space seems to affect the spatial processes at work and alter the resulting long-term representation of space compared to learning in vista space.

1.4.2 Evidence for a Euclidean map

There are a few studies that can indeed be taken as evidence for the formation of Euclidean maps of environmental space. It was shown, for example, that reference frames don't seem to be limited to local vista spaces but instead can spread across and integrate multiple adjacent corridors or streets. Wilson, Wilson, Griffiths and Fox (2007) had people learn four object locations within a simple U-shaped environment consisting of three corridors. In a subsequent pointing task participants performed best when aligned with the initial perspective along the very first corridor compared to being bodily aligned with other orientations. Importantly this pattern was shown independent of the actual position of the participant in the environment during task execution (i.e., standing in the first, second and third corridor). This indicates that a global reference frame covering spatial information from all three corridors, hence, a Euclidean map, was formed for representing this environmental space (see also Meilinger, Frankenstein, Watanabe, Bülthoff, & Hölscher, 2015; Richardson, Montello, & Hegarty, 1999; Tlauka, Carter, Mahlberg, & Wilson, 2011). In line with this, grid cells in rats have been found to realign and form a continuous firing pattern across two adjacent compartments of an experimental box (i.e., two vista spaces) with prolonged experience in the environment (Carpenter, Manson, Jeffery, Burgess, & Barry, 2015). Following these studies, it seems that metric embedding into a continuous Euclidean map may indeed be possible.

It should also be noted that the distortion of distance estimations by opaque barriers and the formation of multiple reference frames that were discussed in the last subchapter are not exclusive to learning an environmental space but can also be found in single rooms. For example, distorting effects of barriers on distance estimations between objects were also found when participants were confronted with transparent barriers (e.g., strings on the floor) that clustered a vista space into regions (e.g., McNamara, 1986; Sherman, Croxton, & Smith, 1979), and multiple distinct reference frames were observed for two object layouts presented within a single room (e.g., Greenauer & Waller, 2010). These results point out commonalities in the representational structure of the two types of space and suggest that spatial memory acquired in vista or environmental space might not be so different after all.

1.4.3 Violations of a Euclidean map?

There are some studies that—at first sight—seem to disprove the formation of a Euclidean map. On a closer look, however, their claims cannot be upheld offhand. For example, distance judgments between objects that are spread across an environmental space were shown to be distorted, namely, mostly overestimated when separated by a wall (e.g., Cohen, Baldwin, & Sherman, 1978; Kosslyn, Pick, & Fariello, 1974; Newcombe & Liben, 1982). However, such biases are insufficient to reject the Euclidean map hypothesis, for even a distorted map can comply with the metric postulates described above (Tobler, 1976). Distorted and incomplete representations have been suggested among others by Downs and Stea (1973) or Golledge, Klatzky and Loomis (1996). For example, based on multi-dimensional scaling techniques Golledge and Spector

(1978) succeeded in generating a distorted but Euclidean map from a city neighborhood based on participants distance estimations. Distortions are immanent in human survey memory. We tend to remember irregular environments as being more parallel and regular, for example, memorizing non-orthogonal turns as 90° turns (e.g., R. W. Byrne, 1979; Moar & Bower, 1983; Tversky, 1981). In cases where participants are still able to draw a coherent map (e.g., Gillner & Mallot, 1998; Tversky, 1981), such findings can again be accounted for by a distorted but Euclidean map. Moar and Bower (1983) had participants judge the angles between triplets of city junctions interconnected by roads, meaning that participants had to judge the pointing angle between B and C when standing at A, the angle between C and A when standing at B and the angle between A and B when standing at C. Thus, the tested triplets of junctions (i.e., three places) formed triangles with mostly non-orthogonal inner angles and sum of inner angles of 180° when regarding the as-the-crow-flies connections between the places. Nevertheless, participants reported mainly 90° angles, which sum up to more than 180° of inner angles for a triplet. This implies a violation of the triangle inequality of metric Euclidean maps. However, as Moar and Bower (1983) noted in the discussion of their results, the local intersections tested usually had an orthogonal layout (i.e., two streets crossing each other orthogonally) and only when a navigator leaves the intersection and navigates along the linking road to the next intersection deviations in the street alignment (i.e., curves) become visible. The memory of the orthogonal local intersections might have been used as a simple heuristic to judge the angles between the neighboring two intersections instead of using the survey relations, which might still be stored in a Euclidean map format. Thus, even though at first sight angular distortions might indicate the violation of the triangle inequality postulate of Euclidean maps, I do not consider these studies persuasive in disproving the Euclidean map approach.

Another Euclidean map postulate that was questioned by a number of findings is the metric postulate of *symmetry*, which states that the distance estimated from A to B must be the same as the estimate from B to A (e.g., McNamara & Diwadkar, 1997). Literature shows that spatial estimates and planning is not always reversely commutative between pairs of places. Different routes were shown to be selected when planning a trip from place A to B compared to planning a trip from B to A (e.g., Stern & Leiser, 2010). Likewise, distance estimations between a pair of places varied depending on which place was set as reference or as target object (e.g., Burroughs & Sadalla, 1979; McNamara & Diwadkar, 1997; Sadalla, Burroughs, & Staplin, 1980). Such asymmetries per se seem to violate the metric postulate of symmetry. However, this dilemma can easily be resolved by proposing that in addition to a Euclidean map another information level is stored which affects survey judgments. One bias model is the category-adjustment model of spatial coding (e.g., Huttenlocher et al., 1991; N Newcombe, Huttenlocher, Sandberg, Lie, & Johnson, 1999), which proposes a fine-grained (conceivably Euclidean) and a categorical level of representation. Categories can, for example, be the four quadrants imposed by vertical and horizontal visual axes of a circle drawn on a piece of paper. Biases emerge during the retrieval of the target location as the fine-grained place information (which can very well be a Euclidean map) interacts with the category level and is distorted towards a category prototype (e.g., the center of a quadrant the target is located in). Thus, asymmetries emerge depending on the prototype that is biasing the estimate and the distance of the target to this prototype. Another slightly different explanation is proposed in the contextual-scaling model (McNamara & Diwadkar, 1997). Every place is assumed to evoke some additional context in working memory when being referenced depending on its salience (e.g., familiarity, functional importance). Depending on the reference object, hence, depending on whether distance has to be estimated from A to B or from B to A, different contexts are activated, scaling the retrieval process accordingly and leading to asymmetries. These two models render asymmetries in distance estimations much less of a violation of metric postulates than originally thought.

Non-hierarchical enriched graph models have the capacity to account for the summarized effects. Distances can be represented distortedly on the local place-to-place level. Enriched graph representations are especially plausible in case of angular distortions that indeed violate the triangle inequality postulate of Euclidean space since distortions are thought to be unique to individual place-to-place metrics and are independent of each other across the entire graph. Since there is no need for forming a globally coherent metric representation no conflict would emerge if angles between three interconnected neighboring places are each represented as 90° angles. Also asymmetries can easily be explained by defining the links between graph nodes as directed (e.g., in following the first walk through the environmental space) and affect retrieval processes accordingly (i.e., slower and more error-prone estimates when retrieving against the directed link) (e.g., Meilinger, 2008) or by assuming not just a single edge that interlinks place nodes but instead two connections associating different travel metrics with each direction (e.g., Mallot & Basten, 2009). But importantly, based on these results it occurs to be invalid to conclude that metric embedding on a global scale (i.e., Euclidean maps) does not exist in human spatial memory.

1.4.4 Evidence for regions and hierarchies

As pointed out earlier already hierarchical representations could combine the enriched graph and Euclidean map representation into one model. Evidence for hierarchical representation of space comes, for example, from studies that have participants learn a number of objects spread across a vista space or from a map. In a study by McNamara, Hardy and Hirtle (1989), first, individual hierarchies were ascertained by deploying an ordered-tree analysis (based on Reitman & Rueter, 1980) on individual participants' object recall protocols. Subsequent distortions in distance estimations between objects and the speed of primed recognition of objects strongly corresponded to these individual spatial hierarchies. This dependency between clusters and performance in primed recognition and survey estimates was also found when clusters were imposed by transparent barriers dividing a vista space containing target objects in four quadrants (McNamara, 1986). Across both studies, general findings were that facilitative priming effects were evoked when priming a target with an object from the same cluster compared to priming with a target from another cluster and that within-cluster distances were underestimated while across-cluster distances were overestimated. But what do we know about clustered environmental space learned from navigation?

In environmental space regional borders and self-imposed regional clusters that circumscribe multiple local vista spaces have been found to affect route decisions, landmark recall, and survey estimates. When having to judge the relative position of remote cities in America (e.g., San Diego in California and Reno in Nevada), participants in a study by Stevens and Coupe (1978) were shown to revert to another level of relational position, namely the one of the two federal states containing the cities (e.g., California is west of Nevada), even if basing the judgement on this higher level conflicted with the actual location of the cities (e.g., San Diego is east of Reno) (see also Friedman & Brown, 2000). One major drawback of this study is the scale of space. Such target stimuli are covering an area that falls into the category of geographic space and have quite probably be learned via a map, hence, represent spatial memory of figural space (Montello, 1993). It is therefore unclear how much such results can be generalized to the spatial knowledge of environmental space that is acquired through navigation alone.

Thankfully, there have been studies delimited to smaller navigable spaces such as city neighborhoods. For example, distorted distance estimations between locations in a multi-neighborhood city area of Ann Arbor were found to be related to subjectively formed clusters (e.g., Hirtle & Jonides, 1985). Similar to the vista space study by McNamara and colleagues (1989) described above these hierarchical clusters—different for each participant—could be unveiled based on participants landmark recall protocols and subsequently correlated with distance judgments. Upon direct comparison within-cluster distances were underestimated while across-cluster distances were overestimated. Such regional distortions could be replicated by Uttal, Friedman, Hand and Warren (2010) on a University campus divided into sub-campuses. They further showed that such cluster biases evolve over prolonged exposure to the environmental space and adduced—based on the category-adjustment model by Huttenlocher (Huttenlocher et al., 1991)—an interaction between a fine-grained and a category level to explain these effects.

These studies done in vista and environmental space are generally accepted as support for hierarchical representations. It should be noted here, however, that this interpretation is not without doubts. First of all, the ordered-tree algorithm used in some of these studies (Reitman & Rueter 1980) is based on and sensitive to whether learned items are recalled in different but systematic orders. It assumes that only after a full cluster has been recalled a new cluster will be selected and recalled, thereby potentially imposing a hierarchical structure where none exists (see also "methodological concerns" in McNamara, Hardy, & Hirtle, 1989). Also, a non-hierarchical representation can be distorted in a way that clusters are formed simply based on the straight-line distances between represented object locations. Objects represented closer to each other form a cluster that is represented distant from other object clusters. Such a non-hierarchical representation, for example in the form of a Euclidean map, might be updated and distorted iteratively (i.e., build up over time) to conform with and exaggerate recurrent clusters a navigator is exposed to. In a neighborhood like the one in Ann Abor this could be "clusters" of places that are regularly visited together during a trip, for example, grocery shopping at the bakery, the butcher and the supermarket or visiting a friend and going to a park and then a bar nearby. Such places might "move" closer together over time. Psychological distances deviating from the real external world can find expression in spatial performance measures. In this case, assuming an additional level of representation for regions and clusters hence is obsolete. This critique is very similar to what was stated above with regards to distorted distance estimations due to opaque barriers: They can still be accounted for by a distorted representation that simply does not mirror external world correctly but was distorted by different factors during encoding.

A promising method to resolve this issue was applied by McNamara (1986), who contrasted participants distance estimates with spatial priming effects in vista space. Non-hierarchical Euclidean map theories would predict that priming effects reflect the erroneously distorted representation, thereby breaking the effect down to simple straight-line distance represented in memory. A linear or exponential decay of priming effects with psychological distance would be expected (e.g., Kosslyn, Ball, & Reiser, 1978). This, however, was not mirrored in participants data, indicating that it is more than just the represented straight-line psychological distance that is driving the priming effect. This is supporting the interpretation that a hierarchical representation was formed. To my knowledge, such a counter-check was not carried out for survey estimates made in environmental clustered space yet.

More solid evidence that regional knowledge is indeed represented in long-term memory and used in spatial tasks comes from Wiener and Mallot (2003) who had participants learn a circular, hexagonal space which was divided into three regions marked by semantic landmark cues (no opaque barriers). In a subsequent wayfinding task, participants had to walk to a goal location in each trial, which could be reached by two equally long routes (clockwise or counter-clockwise around the hexagon to the opposite side). Interestingly, participants preferred the route with the least number of transitions between predefined regions (see also Schick, Halfmann, & Mallot, 2015). Whereas the three regions in the hexagonal space could simply be represented in a distorted way (i.e., without the need to actually represent the regions in long-term memory) that might render the direct path to the other region to be shorter in terms of psychological walking distance compared to walking via the third region, another environment Wiener and Mallot (2003) used did not allow for this alternative explanation. They built a virtual grid field of two rectangles (two regions) arranged directly adjacent to each other and had people navigate to targets within the other region. If regional biases distorted target locations towards the centroids of each region similar distortions would be expected for both regions thereby rendering the path with longer dwelling time in the starting region to be of comparable psychological walking distances as the path with direct transition to the other region (given equivalent path length and complexity in physical terms). Still, they found that participants preferred to directly approach the region containing the target. Thus, potential distortions in memory cannot account for the bias in route decision. Instead, results indicate that regional knowledge was additionally represented in memory. Indeed, representing clusters of places seems to help during navigation. Faster learning and increased efficiency during navigation were found for participants that learned an environmental space clustered into regions compared with no clustering (Wiener, Schnee, & Mallot, 2004).

In sum, these studies support the formation of hierarchical layers in spatial memory. However, they remain relatively uninformative about the format of the regional information that is stored. Effects on route planning found by Wiener and Mallot (2003) can be well accounted for when assuming that the superordinate region- or cluster-levels possess a topological character, thus, representing information about connectivity and containment, but no explicit metrics. For explaining the effects found by Stevens and Coupe (1978) at least some relational properties, for example in the form of verbal labels or metrics, must be stored on the superordinate levels of states (e.g., California is west of Nevada). Distance estimations distorted by clusters/regions (Hirtle & Jonides, 1985; Uttal, Friedman, Hand, & Warren, 2010) could, as described above, be due to distortions in a non-hierarchical representation. Alternatively, a categorical label or a category prototype might bias the estimates during recall. A hierarchy of reference frames (which could also be referred to as hierarchy of Euclidean maps), referring to local reference frames for representing vista spaces and regional reference frames encompassing and integrating information from multiple vista spaces, has been discussed in the literature (e.g., McNamara et al., 2008) but has not yet been experimentally tested for. As mentioned above, findings revealing orientation dependency following a single global orientation (e.g., Richardson et al., 1999; Tlauka et al., 2011; Wilson et al., 2007) suggest that a metric embedding into a global reference system covering a whole region of environmental space is possible.

1.5 Thesis backbone: Assumptions tested

We have seen so far that local vista spaces seem to play a special role in the formation of spatial knowledge. Updating of objects is disturbed when leaving a room (e.g., Wang & Brockmole, 2003a, 2003b), opaque barriers affect remembered straight-line distances between objects (e.g., Cohen, Baldwin, & Sherman, 1978; Kosslyn, Pick, & Fariello, 1974; Newcombe & Liben, 1982), local reference frames each following the geometry of a single corridor are formed (e.g., Meilinger, Riecke, et al., 2013), and when searching for items confusion error occur in the selection of the correct vista space but not in the selection of the appropriate position within the vista space (e.g., Marchette et al., 2017). In sum, navigating an environmental space seems to impose some difficulties integrating spatial information across vista spaces into a coherent metric format.

At the same time, evidence for the formation of a global reference frame covering multiple vista spaces have been found (e.g., Richardson et al., 1999; Tlauka et al., 2011; Wilson et al., 2007) and the ability of grid cells to realign with sufficient experience within a compartmentalized space (e.g., Carpenter et al., 2015). Violation of Euclidean postulates and biases in survey estimates can be met by assuming additional processes involving the context activated by referenced objects (McNamara & Diwadkar, 1997) or additional category representation (e.g., Huttenlocher et al., 1991; N Newcombe, Huttenlocher, Sandberg, Lie, & Johnson, 1999) and by embracing that Euclidean maps do not need to be a perfect reflection of the external world but can be distorted (e.g., Downs & Stea, 1973; Golledge et al., 1996).

Many studies point towards the formation of spatial hierarchies, also when learning object locations in environmental space from navigation only (e.g., Schick et al., 2015; Wiener & Mallot, 2003), but not much is known about the format of regional clusters. Generally, a variety of combinations are feasible for hierarchical representations, for example, an enriched graph on the subordinate level might be formed including local metrics from place to place and regional nodes subsuming multiple places on a superordinate level. These superordinate regional nodes might simply be conceptual labels attached to the places, but they could also be spatially interconnected on the regional level with region-to-region metrics that specify how to get from one region to the other region. A hybrid version of enriched graph representations and Euclidean maps is also conceivable, where places and local metrics are stored on the subordinate level which, however, are globally embedded into a Euclidean map format where each location is assigned coordinate values in a global, mental reference frame (e.g., McNamara et al., 2008).

Taken together, the summarized theories and empirical findings show that there is still experimental work ahead of us to understand how environmental space memory is structured. The Euclidean map approach is widely accepted in part also because it is a very intuitive concept to grasp for the human mind as we are frequently confronted with maps. Much of the literature reviewed so far questions the Euclidean map approach but is not sufficient to disprove it entirely. Many studies cast reasonable doubt on the conception that a compartmentalized environmental space is represented in the same way as locations learned in a clearly circumscribed, fully visible vista space. My working hypothesis is that human survey memory is stored in a non-Euclidean format which might well be enriched by metrics but not globally embedded. If environmental space is not represented within a single, non-hierarchical Euclidean map, the promising alternative is indeed the enriched graph approach or a hierarchical combination. The studies of my thesis are constructed to verify and test the potential characteristics of an enriched graph representation and at the same time collect evidence that might falsify the Euclidean map hypothesis. In the following, I want to deduce several potential effects that should be found in survey tasks performance according to the assumptions made by enriched graph representations. These will be contrasted by predictions based on Euclidean map approaches or the combination of both in a hierarchical fashion. In my thesis I aimed to shed light on the memory units that are stored when representing environmental space (i.e., what is represented in the node of a graph and in which format? Are there local place nodes as well as regional nodes?) and how relations between these memory units are specified (i.e., what's the nature of an edge connecting two place nodes?). Related to the nodes and edges is the question of what metrics are stored. In particular, I explored whether we rely on local place-to-place metrics that are not further corrected for noise and adjusted to form a coherent global geometry. Table 1 lists the effects that I examined in the studies of my doctoral thesis and which will be deduced in the following paragraphs.

Consider again the differences between enriched graph and Euclidean map representations. Enriched graph approaches assume that places form local entities in the graph (i.e., nodes) and survey estimates are constrained to a successive mental activation thereof following the graph connections from one's current location to the target (e.g., Chrastil & Warren, 2014; Meilinger, 2008; Warren et al., 2017). In contrast, having formed an all-encompassing Euclidean map representation of space all spatial information is already represented within a single memory unit and recall should follow a simple readout of coordinates and calculating the direction vector between one's current position and the target using simple trigonometry (e.g., Gallistel, 1990; Gallistel & Cramer, 1996). Based thereon different predictions can be made for the time that is needed to make survey estimates to more or less distant objects. Recalling spatial relations from a Euclidean map should be either done with similar ease for all locations embedded in the map or it might depend on the straight-line (Euclidean) distance between points, which might reflect the time needed for activation to spread across the map or a mental scanning of the map (e.g., Kosslyn et al., 1978). For enriched graph representations the computation time needed for making a survey estimate should be determined by the number of nodes that must be activated along the graph towards the target. Step-by-step each new local memory unit must be added in the computational process, which should render both latency as well as error of the estimate to increase with increasing number of local units along the graph towards the target. At this point I would like to point out that only looking at error accumulation over number of place units can lead to major misinterpretation of the results (e.g., Thorndyke & Hayes-Roth, 1982) as it is not clear whether this error accumulation was already generated during the encoding process and effects simply mirror the precision of memory or whether it is produced during the retrieval of multiple memory entities while executing the survey task. Therefore, it is of importance to look at latency as well which is more strongly associated with the reconstruction process of the memory content (e.g., Pantelides, Kelly, & Avraamides, 2016). In short, while survey estimates based on Euclidean map representation should be bound to the

Table 1. Summary of the effects tested. They cover different concepts that ought to be examined, namely the stored memory units, the nature of their connecting links and the metric information that is represented.

		Examined effects on survey task performance	Study 1	Study 2	Study 3	Study 4
Units/ nodes	Subordinate level	Place-to-place distance effect	●	●		●
		Facilitative alignment effect along local vista space	(●)			●
	Superordinate level	Region-to-region distance effect				●
		Facilitative alignment effect along regional/global orientation	(●)			●
Links/ edges	Order/ Direction	Order of layout reconstruction	●			
		Effect of route direction to the target		●		
Metrics	Local	Biased, globally impossible pointing behavior			●	

Legend: ● Effect was examined in the respective study. (●) Effect was examined but could not be disentangled from a related effect (see summary of studies in the next section).

Euclidean distance between places, recalling from an enriched graph should be bound to the travelled place-to-place distance (i.e., number of corridors/places along the graph). An increase in the distance unit should lead to a decrease in performance. The same logic can be applied for superordinate levels in a hierarchical representation. For example, if distinct regional memory units are formed that subsume a subset of places (whether in a metrically embedded Euclidean map format or not) participants should take more time to point across regional boundaries compared to pointing within regional boundaries, as the former requires the activation of a new memory unit. Thus, in the case of such intermediate hierarchical levels, region-to-region effects on latency are assumed. I tested these distance effects in study 1, 2 and 4.

How are relations between place units specified in memory? Corresponding to the arguments made in the previous paragraph, inherent in the concept of the Euclidean metric map is the idea that such a representation must possess an allocentric all-encompassing format that relates "every point on the map […] to every other point" (p.166, Nadel, 2013). Thus—within the limits of distortions of the psychological distances between learned object locations—the Euclidean map format should be abstracted from the actual egocentric learning experience in the environmental space. This implies that no order effects based on the sequential encounter during learning should prejudice the process of estimating survey relations. In contrast, enriched graph models are sequential in nature, specifying the walked translations between neighboring places. Therefore, also memory should be accessed following along the graph from one's current location to the target location. Additionally, enriched graph models often assume that the links connecting two local places are directed, specifying unique travel instructions to get from the current place to the other but not the other way around (e.g., Mallot & Basten, 2009; Meilinger, 2008). Based on these assumptions we would predict that the experienced successions of spatial information across an environmental space should be directly specified in graph memory and affect computations based thereon. For example, layout reconstruction should follow the experienced learning order and/or survey estimates should be facilitated in the direction the route was learned. These order and direction effects were tested in study 1 and 2.

How are metrics of environmental spaces stored? Enriched graph representations assume that local place-to-place metrics are memorized and reactivated successively to estimate survey relations on the fly when needed. Importantly, these local metrics do not need to be globally consistent. Instead, each local metric is independent of the other local metrics. This means, for example, that a triangle of three places can be represented without sensing a conflict even though the sum of inner angles that are stored as part of the local metrics might be subject to erroneous encoding and exceed or undercut 180°. Euclidean maps, on the other hand, need to obey metric postulates. Thus, the stored metrics must be adjusted to achieve global embedding and to assign unique coordinates in a mental coordinate system. Normally distortions in represented metrics arise from our flawed path integration system (e.g., Loomis et al., 1993; H. Zhao & Warren, 2015; M. Zhao & Warren, 2015). We amplified this by testing how participants handle local place-to-place metrics that induce a massive conflict with Euclidean metric postulates. We had participants learn a non-Euclidean environment where the local place-to-place metrics did not match up on a global scale. Gaps in physical space were covered by non-overt teleportation in virtual space. Enriched graph representations would be able to preserve the local place-to-place metrics as experienced without evoking any conflict in the representation. Euclidean metric embedding, however, requires adjustments and corrections to be made to all experienced metrics in order to account for the teleportation gaps and to form a globally consistent representation. Whether local place-to-place metrics are used as experienced or whether global embedding occurs was tested in study 3.

The enriched graph representation of Meilinger (2008) assumes that local vista spaces are represented as local reference frames, thus, small Euclidean maps limited to the circumscribed area of a corridor or room. These local reference frames are connected in a graph network, but not further embedded on a higher level. Opposed to this, the Euclidean map approach postulates that a single map (i.e., global reference frame) can embed all locations encountered in environmental space. In my thesis, I also wanted to address the possibility that not only single layer representations are formed, but also hierarchical representations (e.g., Stevens & Coupe, 1978; Wiener & Mallot, 2003), which might even combine graph and map approaches into a single model. Therefore, I confronted participants with an environmental space consisting of multiple corridors that are divided into two saliently defined, interconnected regions. My approach basically translates to the formation of multiple Euclidean maps on multiple levels of the hierarchy: Multiple local coordinate systems on the lowest level one for each corridor encountered, two distinct coordinate systems on the intermediate regional level entailing the spatial information of the local corridors of a region and potentially a single global coordinate system comprising the entire space learned. Such a metric embedding into Euclidean maps on multiple levels should manifest in orientation dependencies of survey estimates that follow local, regional and global main orientations. Bodily alignment with the main orientation of a reference frame should yield better survey performance compared to being aligned otherwise. Alignment effects of orientation-dependent spatial memory were tested in study 1 and study 4.

Having derived now all assumptions that I tested in my doctoral thesis another relevant question remains, namely, whether such effects are specific to environmental space knowledge or whether they can also be found in vista space? This question actually was the starting point of my examination. In my first study, I tested whether two of the assumptions deduced from enriched graph representations can be found in spatial memory acquired in a single vista space as well. I directly compared the formed survey representation for the same object arrangement that was either learned fully visible within a single room (vista space) or separated by walls that made up multiple interconnected corridors (environmental space)[3]. By keeping the Euclidean distance and relations of the object layout exactly the same across conditions I prevented any confounding effects of environmental scale (as it is often the case when comparing different studies investigating either vista or environmental space, where environmental spaces are often much larger). Additionally, I aimed to entangle the effects of opaque barriers (walls), physical movement through space and the successive encounter of objects which typically accompany navigation through environmental space and contrasts with the natural experience within a vista space. Specifically, I tested whether survey estimates based on a vista space representation are also bound to unit-by-unit distance effects and whether the recall of spatial relations is constrained to a predefined order set by the learning experience. Both effects would be expected by enriched graph representations of environmental space.

In sum, in the studies of my doctoral thesis I examined unit-to-unit distance effects, order and direction effects, local bias effects and alignment effects to explore the long-term memory structure of environmental space representations. In four studies I covered the characteristics of local, regional and global map memory units, their connecting links and stored metrics and their potential hierarchical structuring. I sought to provide evidence that environmental space—in contrast to vista space—is stored in a clustered graph format and not embedded in a globally consistent Euclidean map.

---

[3] Vista space studies usually test what they call *object-to-object relations* (e.g., Avraamides & Kelly, 2010; Mou & McNamara, 2002; Yamamoto & Shelton, 2009). Object-to-object relations formed in environmental space are subsumed under the term *survey knowledge* (e.g., Siegel & White, 1975). For the sake of convenience, I use the term *survey knowledge* for both the vista and the environmental space condition.

2 Summary of studies

In this section, the four studies of my doctoral project will be briefly introduced and summarized individually. A broader discussion encompassing all results will be made in the General Discussion section. The focal point will be to elaborate upon the predictions worked out in the last section. Nevertheless, each study also covers some novel unique aspects not yet introduced or discussed but still worth noting. Each study in its entirety and including figures can be perused in section 7 “Full publications and manuscripts” at the end of this thesis for an in-depth comprehension.

All spaces used were virtual environments especially built for experimental purposes and contained virtual target objects. The experimental setup of study 1, 3 and 4 consisted of a large tracking space participants physically walked through while their head movements were tracked by cameras. A head-mounted display connected to a computer rendered a real-time egocentric view of the virtual environment according to participants body and head movements. Study 2 covered a large virtual environment exceeding the walkable area of the tracking hall. Thus, an omnidirectional treadmill was used that moved participants back to the center of the treadmill with each step they took. Survey estimates were executed either via a 360° movable joystick or via pressing a button on handheld controller device while facing the estimated direction of the target. Both setups enabled free physical movement through the virtual environments and thereby provide proprioceptive and vestibular feedback (existent also for the omnidirectional treadmill at the beginning of movement and during and shortly after change of heading as the centering of the treadmill was accompanied with some latency), efference copies and visual cues such as stereo vision, optic flow, motion parallax and occlusion. The importance of body-based senses for acquiring accurate information about spatial metrics has been supported in many studies (e.g., Klatzky et al., 1998; Ruddle & Lessels, 2009). Thus, the setups used in the current studies are particularly suitable to examine how environmental space long-term memory is structured and to allow generalizability to spatial learning in real, physical space.

2.1 Study 1: Vista vs. environmental space

2.1.1 Research question

The aim of study 1 was to ascertain how survey knowledge acquired within a vista space (also often referred to as object-to-object relations, e.g., Avraamides & Kelly, 2010; Mou & McNamara, 2002; Yamamoto & Shelton, 2009) differs from survey knowledge acquired in environmental space. As indicated by the literature reviewed in the introduction differences in the structure of survey memory can be expected, however, adequate comparability between the two types of space in terms of memory load and difficulty was never ensured so far (e.g., Brockmole & Wang, 2002, 2003; Kosslyn et al., 1974; McNamara, 1986; Newcombe & Liben, 1982). Therefore, in my study, the same number of objects had to be learned in both vista and environmental space covering a comparable area in size. In a second step, the origin of potential differences between memory for vista or environmental spaces was examined. In contrast to vista space, environmental spaces are concomitant with the presence of opaque barriers compartmentalizing space into multiple local vista spaces, the need to physically move through these local compartments for full coverage and the successive nature of object encounter along the path through the environment. Indeed, both opaque and transparent barriers were found to effect distance estimates between objects (e.g., McNamara, 1986), rendering the other two factors to be of potential relevance for structuring survey knowledge as well.

Walking enriches the learning experience by visual and proprioceptive inputs (including a more direct experience of some inter-object distances in the case of our experiment) and multiple perspectives onto the object layout. Proprioceptive input alone (blindfolded walking across locations of an object layout) was shown to be sufficient to yield a facilitative reference orientation independent from an additional reference orientation set by visual learning (e.g., Yamamoto & Shelton, 2005, 2007). Regarding multiple views within an environment it was found that despite the importance of the first view experienced within a vista space for setting the reference frame orientation (e.g., Kelly & McNamara, 2008; Rieser, 1989), also bodily alignment with salient layout intrinsic or extrinsic cues later during learning was shown to determine the reference frame orientation (e.g., Kelly & McNamara, 2008; Shelton & McNamara, 2001; Valiquette & McNamara, 2007).

Successive encounter of objects within an environmental space does not only provide strong spatiotemporal cues along an order predefined by the path, but additionally affords simultaneous visual access to only a subset of relevant objects (e.g., distance and relative position between object A and B located within one local environment of an environmental space) while preventing this direct input for other objects (e.g., object B in first vista space and object C in the neighboring vista space). Thus, some spatial relations must be inferred across multiple visual experiences.

For Experiment 1 of the first study comparable conditions for learning an object layout either in vista or environmental space were created in virtual reality. Seven objects arranged on the floor in an incomplete 3x3-grid with bilateral symmetry (closest row three objects, middle row three objects, farthest row one object in the middle) were presented to the participants in the middle of a virtual room (vista space group) or spread across multiple parallel corridors (environmental space group). In the vista space the object layout was aligned with the room geometry (rectangle with longer walls along the visual perspective of the participant and the elongated midline of the object layout). For the environmental space group, the object layout was identical to the vista space layout (same Euclidean distance and relative direction between the objects). To create the environmental space virtual walls were erected that compartmentalized the space forming four parallel, interconnected corridors (each containing one to two objects) obliquely aligned to the global object layout and the room of the vista space condition. Permanent objects at the wall (e.g., window, plant) provided an overall orientation within the local environmen*t*(s). The vista space group learned the object layout while standing at a predefined vantage point, looking along the midline of the object layout. The environmental space group started off from the same location relative to the layout but walked along all corridors multiple times to encounter all objects. After learning the object layout was removed from the scene and two survey tasks had to be solved, a visual pointing and a layout reconstruction task. In each visual pointing trial participants were teleported to different locations within the virtual environment standing right on top of one of the seven object locations, being bodily aligned with one of eight orientations. These orientations were arbitrarily labelled with respect to the underlying object layout, 0° orientation referring to being bodily aligned with the main axis of the layout (midline towards the single object in the last row) and from there 45°, 90°, 135° rotated to the left or right or contra aligned, labelled 180°. Participants had to point with a joystick to one of the remaining six target objects, which could be either located within one's current corridor, in a neighboring corridor or two or three corridors away, rendering the factor corridor distance. In the layout reconstruction task objects were presented in random order in a row in front of the participants and the arrangement had to be reconstructed from memory.

We were interested in three possible effects that would distinguish between the two types of space. If the compartmentalization of space (environmental space) yields a likewise compartmentalized representation (e.g., a network of reference frames; Meilinger, 2008) that is not isolated from the egocentric experience during learning, we first predicted that the layout reconstruction of the environmental space group should follow the order of the first encounter through the environment. Second, during the visual pointing the individual local memory units should be activated successively on the fly to incrementally calculate the target direction relative to one's current position. Therefore, the more local memory units (i.e., corridors) are residing between current and target position, the higher the computational effort, and thus, the higher the pointing latency. Finally, the intrinsic cue of the global layout should be of negligible relevance for setting the reference frame orientation as it is not directly perceived during learning. Instead the orientation of the local corridors (oblique to salient layout-intrinsic axis) should set the reference frame orientation. In contrast, if survey estimates are based on a global, all-encompassing, allocentric Euclidean map, performance in the environmental space should be similar (or at least approximate) performance displayed in the vista space group. Survey performance should be abstracted from the order of learning experience (no order effect in reconstruction task) and all information from the local compartments should be integrated into a single representation unit, therefore, being accessible with similar ease (no corridor distance effect on pointing latency). Also, global cues from the entire layout (main axis of layout) might get more weight in setting the orientation of the reference frame, thereby shifting the alignment of the reference frame to the salient layout axis.

In Experiment 2 of study 1, we ran three additional conditions that emulated the learning experience of the environmental space in a vista space setting. One group of participants walked the exact path participants in the environmental space group traveled, but with full view of the entire layout within a single room (*movement-simultaneous objects*). Another group viewed the room from a single vantage point without moving, but the object layout was presented to them successively in the same order and grouping as experienced in the environmental space condition (*static-successive objects*). The last group moved along the predefined environmental space path through the vista space room while subsets of objects were presented to them depending on their current location (*movement-successive objects*). Implementing movement through space and successive object encounter without compartmentalization of space along opaque barriers allowed us to examine which component is the driving factor for potential differences between vista and environmental space survey knowledge.

2.1.2 Main results

Learning in the compartmentalized environmental space led to (a) an effect of corridor distance on pointing latency (i.e., increase in latency the farther away the target was located in terms of travelled corridors), (b) best pointing performance (lowest error and latency) when aligned with the local geometry of a corridor (oblique to global layout geometry) and (c) layout reproduction following the order of first encounter through the corridors (in contrast to following an order along rows/columns of the object layout or the random object presentation at the beginning of the reconstruction task). These patterns differed significantly from the performance displayed by the vista space group. Here, pointing latency was neither dependent on corridor distance (albeit irrelevant for the raw vista space condition, as no corridors were experienced) nor on Euclidean distance between current and target location, best pointing was observed for body orientations along the layout geometry (consistent with first view and room geometry) and the layout was reconstructed in an order either following a sequential work-through of the row/column of the layout or the random presentation. Neither of the hybrid conditions in the vista space room of Experiment 2 emulating walking and successive object encounter typical for environmental space yielded an effect of (invisible) "corridor" distance nor did they render best pointing performance when participants were obliquely aligned to the global layout and room geometry. Patterns significantly differed from the environmental space condition, but instead resembled those found in the original vista space condition. Only for the layout reconstruction task mediocre relocation preferences along the order of environmental space learning were found for the static-successive objects and movement-simultaneous objects groups. Yet, these medium correlations differed significantly from the high correlation found for the environmental space group, except for the static-successive objects group.

2.1.3 Summary of study 1

Albeit learning the exact same object layout survey memory acquired in vista space differed fundamentally from that acquired in environmental space. Retrieving survey information from an environmental space was bound to the functional distance (corridor distance effect on pointing latency, which is not driven by Euclidean distance) and order (reconstruction order) exposed to during learning. Thus, no abstraction from the egocentric learning experience was achieved in environmental space and not all targets were accessible with similar ease. This contrasts the vista space condition where layout reconstruction was much more flexible and distance to the target (neither invisible corridor distance nor Euclidean distance) had no effect on the time needed for estimating the target direction. Additionally, for the same object layout obliquely aligned reference frames seem to have formed in both types of space, each reference frame following the geometry of the local environment (corridors or single room). Neither movement along a similar path, successive object presentation nor their combination implemented in vista space yielded similar effects as found in the environmental space condition. Only order effects during layout reconstruction can at least be partly accounted for by these factors. Nevertheless, the main factor dissociating survey memory for vista and environmental space occurs to be the presence of opaque barriers.

The local geometry of the individual corridors in the environmental space seemed to have off-set the reference frame alignment compared to the vista space condition. Yet such a result is inconclusive on whether the found orientation dependencies mirror the use of a global reference frame encompassing all seven objects (e.g., Meilinger, Frankenstein, Watanabe, Bülthoff, & Hölscher, 2015; Tlauka, Carter, Mahlberg, & Wilson, 2011; Wilson, Wilson, Griffiths, & Fox, 2007) or multiple local reference frames, each limited to an individual corridor (e.g., Meilinger et al., 2014; Werner & Schmidt, 1999). A similar ambiguity with respect to these two alternatives can be identified for previous studies concentrating on the accuracy of spatial memory, for example, how well psychological distance resembles physical distance. For example, findings that opaque barriers (but also transparent barriers, whether walked across or not) bias distance estimations between targets (e.g., McNamara, 1986) can both be explained by a strong distortion of a global cognitive map due to error accumulation during learning or by the successive recall process of multiple sub-maps that biases and increasingly distort the estimates (e.g., Fujita, Klatzky, Loomis, & Golledge, 1993; Meilinger, 2008). However, evidencing corridor distance effects on pointing latency as done in this study yields an important insight into the structure of survey knowledge. It suggests that the representation itself is compartmentalized into subunits, for example, one unit for each visited vista space. The process of computing survey estimates seems to rely on these local memory units and is bound to the successive, ordered recall thereof. Accounting for such distance effect based on a globally integrated Euclidean map seems challenging (but this will be addressed in the General Discussion of this thesis). Instead, our results seem to support enriched graph theories.

2.2 Study 2: Routes embedded in survey knowledge

2.2.1 Research question

Study 1 already indicated that vista space subunits have been formed and that survey knowledge is not abstracted from the order of learning. These memory characteristics seem to be strongly influenced by barriers occluding the view upon spatial locations that must be related to one another. Spatio-temporal aspects defined by the egocentric experience within environmental spaces are typically associated with route knowledge. For example, results by Strickrodt, O'Malley and Wiener (2015) suggest that routes are not navigated based on simple stimulus-response association, meaning, that the appropriate decision (left, right, straight) follows as a response to a landmark at an intersection. Instead, the representation seems to be much more complex and interlinked. A tight coupling of the succession of landmarks along a route each associated with a decision to make allows for recalling the correct route decision when faced with ambiguous intersections. For example, when confronted with an intersection containing a non-unique landmark (i.e., a similar landmark was already encountered somewhere else along the route) the identity of the preceding landmark as well as the route decision made there was shown to be sufficient to retrieve the correct route decision on the next intersection (Strickrodt et al., 2015). Hence, single decision points are integrated into a richer route representation best described by stimulus-response-stimulus associations (e.g., O'Keefe & Nadel, 1978; Schinazi & Epstein, 2010; Strickrodt et al., 2015; Wiener, Kmecova, & de Condappa, 2012). Correspondingly, these links between connected places of a route were shown to be directed. Primed recognition experiments repeatedly evidenced what is now known as the *route direction effect*. Being primed by a landmark that preceded the target object during learning speeds up recognition of the target compared to being primed with a landmark that succeeded the target along the route (e.g., Janzen, 2006; Schinazi & Epstein, 2010; Schweizer, Herrmann, Janzen, & Katz, 1998).

The directed graph model proposed by Meilinger (2008) assumes that also survey estimates are constructed based on local memory units connected by directed links (see also Mallot & Basten, 2008). Thus, having learned an environmental space along a predefined route subsequent survey estimates that are made to targets located route forward towards the end of the route should be made faster than estimating survey relations to a target located route backward towards the start of the route. In contrast, following a stern definition of a Euclidean map would preempt the existence of route direction effects in survey estimates altogether as allocentric configurational knowledge is typically thought to be uncoupled from the order of learning. Therefore, in study 2 we set out to examine whether this so-called *route direction effect* is also immanent in survey knowledge. After learning eight locations along a virtual route at least six times in the same direction (always from start to end), while wearing a head-mounted display and walking on an omnidirectional treadmill, participants were again teleported randomly to different locations and had to point to targets located route-forward or route-backward within the environmental space. This approach differs from the free layout reconstruction of study 1, as now the to-be-recalled spatial information of one's current location and the target is predefined, allowing to observe the processing time for survey estimates in or against the learned route direction. In addition to manipulating the route direction of the target participants always had to point from one standpoint to multiple related targets one after another, more precisely, the subsequent target within a chunk of trials was always a direct neighbor to the preceding target. Within a chunk of trials either all targets lying towards the start of the route or all targets lying towards the end of the route relative to one's current location had to be pointed to successively. We balanced whether this was done in a sequence away from ones current location (i.e., first pointing to one's direct neighbor, then the following target etc., until reaching the target at the start/end of the route) or in a sequence starting at the outer part of the route and successively querying targets in a sequence towards one's current position (i.e., first pointing to the start/end of the route, then to the second/second last target along the route etc. until having to point to one's direct neighbor). Testing such chunks of interrelated trials allowed us to investigate whether subsequent survey estimates are based on previous estimates or whether they are of such transient nature and independent that every new estimate must be made from scratch again. While a simple read-out of coordinates from a cognitive map reference system would predict comparable performance for first and later pointings within a chunk (i.e., independence), incremental construction models based on a graph structure allow to base subsequent estimates on previous estimates, which might ease the computational effort. Furthermore, since pointings had to be made from all places along the route different *place-to-place distances* (i.e., number of places along the route between current and target location) were queried across and within a chunk of trials, allowing to analyze whether the distance effects found in study 1 can be replicated.

2.2.2 Main results

With regards to latency, we found a route direction effect on survey estimates for the first target within a chunk of trials. Participants pointed faster to targets lying in route direction than to targets lying against route direction towards the start. Additionally, latency correlated with the leg distance to the target, indicating that the time needed to estimate the direction increases with increasing route distance to the target. Due to the almost circular nature of the route (i.e., the end of the route bends towards the start again thereby minimizing the Euclidean distance), this leg distance effect cannot be explained by an effect of Euclidean straight-line distance. Both the route direction effect and the place-to-place distance effect on latency disappeared for later pointings within a chunk of trials. The accuracy pattern (absolute pointing error) largely mirrors the latency pattern of results—for first pointings error is lower when the target resides route forward towards the end and error increases with increasing leg distance to the target—with the only difference that the distance effect is still present for later pointings within a chunk of trials. Overall participants pointed slower but also more accurate in their first pointing compared to later pointings.

2.2.2 Summary of study 2

The observed route direction and effects on pointing latency for first pointings within a chunk of trials support graph theories assuming route forwards encoding in the form of a directed graph structure. As survey estimates are presumed to be directly constructed from that directed graph the integration of local memory units towards the end of the route should be sped-up, while integration towards the start of the route should be slowed down. Our results, therefore, suggest that the route direction effect found in landmark recognition tasks (e.g., Schweizer et al., 1998) can be generalized to survey knowledge. We observed route direction effects both on latency and error. This supports but also extends previous findings where the route direction of the target was found to effect pointing accuracy after participants learned a route in one (Moar & Carleton, 1982) or both directions (Meilinger, Henson, Rebane, Bülthoff, & Mallot, 2018). Evidencing asymmetric effects on accuracy after enabling participants to experience both directions (Meilinger et al., 2018) allowed for an alternative explanation, namely, that two separate distorted Euclidean maps for each walking direction were formed and selected respectively depending on the location of the target (in or against route direction). However, as in our study the environment was experience only in one direction and effects were found both in error and latency the observed asymmetry must be immanent in the format of a single representation thereby influencing the spatial processing of survey estimates. The observed distance effect replicated findings from study 1 and further supports the assumption that survey estimates are done incrementally based on a graph structure, successively activating the local memory unit along the experienced order.

Later pointings that followed the first estimate were faster but also more error-prone than the first pointings. Interestingly both the distance and the route direction effect were only evident in the first trials of a chunk of interrelated trials. Later pointings were independent of whether the target was located route forward or route backward relative to one's current position, and for latency also independent of the leg distance to the target. Bearing in mind that within a chunk of trials successive pointings were always just one intersection away from the previous estimate (i.e., target-to-target distance is 1) our results suggest that participants did not repeat the incremental process of integrating all intersections between their current location and the new target again (which should have resulted in distance effect on latency for later pointings as well), but instead only added or subtracted the single segment between the old and the new target to their previous estimate[4]. This makes the estimate on average faster for later pointings but also increases error. That is because when building upon previous estimates the number of estimates across the chunk of trial adds up and increases error for every mental processing step that is made. The distance effect on pointing error which sustained for later pointings within a chunk of trials can be explained by error accumulation during learning. Assuming a roughly constant random error during encoding, integration across larger distances during learning will aggregate larger errors. Thus, distance effects on error for first and later trials within a chunk may simply reflect the accuracy of the survey representation.

Taken together our results suggest that estimates are done incrementally along an enriched graph structure that incorporated the directedness of the learning experience. Each estimate is generally transient because at the beginning of every new chunk the process had to be reiterated, but within a short timescale the estimate can be maintained in working memory to serve as a ba-

[4] Note that the route direction effect for later pointings did not invert in chunks following an order starting at the outer ends of the route towards ones' current location. In these cases, the route direction along with the first estimate is constructed is invers to the route direction along which subsequent targets might be constructed if later estimates were based on estimates of previous targets. This inversion, however, is not reflected in participants performance in the form of a reversed route direction effect. Thus, the role of route direction for later pointing is not yet fully clear. Potentially participants accessed the previously constructed mental model parts still present in working memory.

sis for subsequent survey estimates. An all-at-once read-out from a Euclidean map could not account for these effects. Instead, utilizing a Euclidean map would predict comparable performance for first and later pointings within a chunk, no route direction, and no place-to-place distance effect.

2.3 Study 3: Learning a non-Euclidean environment

2.3.1 Research question

Study 3 aimed to investigate whether survey knowledge is embedded into a globally consistent format, as postulated by Euclidean map approaches. More precisely, we examined whether pointing patterns conform with the metric postulate of positivity (see Beals et al., 1968; McNamara & Diwadkar, 1997; Warren et al., 2017). It follows the assumption that a Euclidean map representation should be stored in a coherent global format and it defines that every place is assigned unique coordinate values which cannot be occupied by other places. A coordinate-free but still globally consistent version of the Euclidean map idea has been explicated for example by Mallot and Basten (2009) (see also Sholl, 2001; Easton & Sholl, 1995; Sholl & Nolin 1997). Local metric information can be checked and optimized for global consistency by triangulation without the need to assign coordinates. For both the coordinate-free and the Cartesian coordinate version of the Euclidean map an embedding into a globally consistent metric format is assumed. This infers that survey estimates from one place to a fixed other place should result in identical responses. In other words, pointing from A to B should yield a uniform pointing direction (within the normal range of variability). Enriched graph representations do not presuppose global consistency. Instead, local metrics can be independently distorted. To test which approach described the format of survey knowledge best we decided to have participants learn an impossible, non-Euclidean space in study 3 and ascertain how their spatial memory system deals with this situation.

Virtual impossible worlds have been used previously in the spatial cognition literature. Typically, visually seamless wormholes are used which upon contact teleport participants to different locations in the environment. This leads to traveled paths within physical space that overlap in an impossible manner and to shifts of virtual places to different locations in the external physical world every time the wormhole is passed. Whether learned via a 2D projection or via walking participants, were found to be well able to find the shortest routes to objects located within the impossible environment (e.g., Ruddle, Howes, Payne, & Jones, 2000; Warren et al., 2017; Zetzsche, Wolter, Galbraith, & Schill, 2009), indicating that local place-to-place connections are stored and used for route planning. Kluss, Marsh, Zetzsche and Schill (2015) had participants learn relatively simple virtual environments, for example, three corridors forming a triangle. While the possible version of the triangle possessed a sum of inner angle of 180° the impossible version used widened up individual angles of 90° that summed up the inner angle to 270°. Examining participants turning angles during blindfolded re-traversing of the path showed that the sum of turned angles was around 270° indicating the representation and use of local metrics that do not match up to a globally consistent triangle. Unfortunately, no difference could be found from the behavior in the possible triangle environment due to a high variability in responses rendering the result less persuasive in its interpretation that no global embedding into a Euclidean map occurred. A more complex environment was used in a study by (Warren and colleagues (2017). Global inconsistencies of the local metrics that were traveled occurred between the walking trajectories from object A to B via the constant midpoint of the maze or traveling there via a route crossing a wormhole. Straight line directional estimates revealed that—compared to the possible maze group that learned a natural scenario of the maze—near-wormhole objects in the impossible maze were represented as ripped apart from other close-by object locations and biased strongly towards their expected wormhole locations. However, since testable predictions for a potential global embedding of the impossible space were missing in this study results remain ambiguous on whether the observed biases can also be explained by a highly distorted but an overall globally consistent Euclidean map. Indeed, modeling the distortions induced by another wormhole maze including the rotation of local corridors and the displacement of the target object relative to the ground truth (no wormhole condition) as parameters to the model was found to provide a good explanation of participants pointing behavior (e.g., Muryy & Glennerster, 2018).

In study 3 we wanted to take remedial action to resolve this ambiguity. Participants either learned a possible or an impossible, complex, seven-corridor environment. Both environments were circular, thus, the last corridor was directly connected to the first corridor, enabling a continuous walk through the environment for multiple laps. Each corridor contained an object identifying the place that must be learned. In the possible maze the seven objects were positioned at the vertices of a regular heptagon. The impossible maze was a disjointed, widened-up version of the possible environment where the same seven places were mapped onto seven adjacent vertices of a decagon, leaving two vertices unfilled thereby creating a gap. Thus, after walking one round, reentering the starting corridor, participants ended up at a position in real, physical space a few meters distant from their starting location. Here the virtual environment was rotated visually seamless, matching up a duplicate corridor with participants' current position and continuous exploration could carry on. Each round the local corridor-to-corridor metrics remained constant although the space was not matching up on a global scale. Path length and straight-line distance between direct neighbors were the same across both environments.

Subsequently, participants had to solve a pointing task. They were teleported to different objects and had to point towards four other objects one-by-one. The sequence of targets within a chunk of trials was predefined, following either a clockwise order around the circular environment (starting with the nearest neighbor to participants left, followed by the next neighbor and so on) or a counterclockwise order (starting with the nearest neighbor to participants right, etc.), resulting in the factor *relative corridor distance* (1-4) and *order of target sequence* (clockwise vs. counterclockwise). The last two of the four targets within a chunk queried in clockwise order overlapped with the last two targets queried when standing at the same location but following a counterclockwise order. Hence, these were the trials were participants actually had to point from the same position to the same target but being either primed along the clockwise or the counterclockwise direction. Considering a graph representation, by manipulating the target order we established which graph nodes and edges are activated (clockwise vs. counterclockwise along the circular graph) and hence which local place metrics are used to estimate the target directions. This lends from the logic of study 2, where we found that querying neighboring places in succession seems to operate like a single cycle of related estimates that sustain in working memory, each succeeding one based on the preceding one (until a new teleportation disrupts this process and requires the operation to start anew. If estimates are based on an enriched graph representation local place-to-place metrics should be systematically biased and globally inconsistent in the impossible maze group. In contrast, if a Euclidean map representation is formed local metrics from the clockwise and counterclockwise direction should be embedded in a common coordinate system. Thus, the gap introduced by spreading seven objects across seven of ten decagon vertices should be closed. When averaging across all seven standpoints from which a participant had to face other targets a roughly even spread of the objects around a circle should be approached. Hence, in the case of global embedding participants should point along a heptagon layout similar to that of the possible maze group.

A possible mechanism that could support global metric embedding of those local place-to-place metrics is the constant updating of the locations of visited places relative to one's changing position through an environmental space and a recalibration and correction thereof when reaching a known location (e.g., getting back to the first corridor after one walk through the maze). Such a mechanism was proposed, for example by Wang (2016), who specified that a navigator might carry a set of vectors each pointing to visited places. Upon view of one of the known places the predicted location and the actual location in view can be compared and used to synchronize the set of vectors. A similar approach has been proposed and simulated for robot navigation by Hübner and Mallot (2007). These proposals have been made to compensate for the imperfect path integration systems in robots and in humans (e.g., Loomis et al., 1993; Zhao & Warren, 2015a, 2015b) that involve accumulating of error with every step taken and every angle turned (e.g., Fujita, Klatzky, Loomis, & Golledge, 1993). However, they are just as well suitable to explain how our participants might be able to make global sense of our impossible world. In short, the question I asked in study 3 was whether participants independent of the biased direction actually point towards the same place when indeed asked to do so?

2.3.2 Main results

The possible maze group was not affected by the biased direction within a chunk of trials (clockwise vs. counterclockwise), instead their pointing pattern was accurate and not significantly different from the underlying heptagonal layout of the objects. In contrast, an effect of the order of target sequence was found for the impossible maze group. When having to point to the exact same target objects participants actually showed a significant leftward bias when pointing successively along a clockwise target sequence and a significant rightward bias when following a counterclockwise target sequence. In other words, their direction estimates towards the same targets differed significantly, each showing a strong outward bias. We then contrasted the groups pointing patterns over all four relative corridor distances queried within a chunk, both relative to the baseline prediction of the possible maze heptagon (likewise the prediction for global embedding) and relative to a linearly increasing outward bias across corridor distance that is predicted when storing the experienced local place-to-place metrics without global embedding. First, the patterns of the two groups differed significantly. Second, the outward bias of the impossible maze group (i.e., deviation from the Euclidean map prediction) increased with increasing corridor distance to the target. This was reflected in post-hoc t-tests evidencing significant differences in outward error across queried distances and reflected in a positive slope which was averaged from the individual slopes taken from independent linear regression run on the full data of every single participant over the four relative corridor distances. Third, the error pattern of the impossible maze group was best described by the systematic increasing outward bias predicted by the enriched graph model. Pointing directions were close to the enriched graph prediction (i.e., decagon) for the first three targets. Only the last target within a chunk yielded a constant error that swung off from the enriched graph prediction towards the global embedding prediction. Neither the enriched graph nor the Euclidean map prediction could account for this "flattening" pattern. Importantly, the within-subject variable error (*SD*) was the same for both groups and increased steadily over the four targets in both the possible and the impossible maze group. Additional the latency pattern across the four targets was highly comparable. Especially, no latency difference was observed between the estimate to the second, third and the fourth target. Both *SD* and latency patterns indicate that no switch in strategy or change in estimation process occurred for relative corridor distance four. Further, they suggest that the same mental processes underly survey estimates in both groups.

Nearly half of the participants in the impossible maze group (but also 22% in the possible maze group!) reported noticing something unusual about the environment (e.g., "I feel the environment might not be a circle"). Still, their error pattern did not significantly differ from those who did not notice anything. Both the noticer and not-noticer pointed out different directions to the same target and yielded error patterns that followed the impossible local place-to-place metrics.

2.3.3 Summary of study 3

Study 3 clearly showed that the impossible maze was not brought into a coherent, globally consistent memory format as predicted by the Euclidean map theory. Instead, pointing patterns corresponded nearly perfectly with the local place-to-place metrics that have been experienced, causing the participants in the impossible space group to point out different directions albeit being queried the same object. Maze architectures for the possible and impossible maze group showed high similarity in most terms. The same number of objects located in the same number of corridors which are connected in a highly comparable circular manner with identical neighbor-to-neighbor straight line distances but with only slight differences in connecting corridor angles and individual corridors length but overall covering the same path length had to be learned in both groups. The only difference was, that the local place-to-place metrics in the impossible maze did not match up on a global scale but instead widened up the baseline object layout of the possible maze (regular heptagon) to fit on seven vertices of a regular decagon leaving a gap in physical space. Indeed, our results show how sensitive humans can be to the local metrics encountered and how precise they can be stored and recalled.

No global embedding seemed to have occurred. As pointed out above, updating and optimization upon visual reference could be a potential mechanism enabling global embedding (e.g., Hübner & Mallot, 2007; Wang, 2016). Research suggests that up to six objects can be updated in parallel without losses during blindfolded movement (Hodgson & Waller, 2007). Thus, with seven objects in study 3, we might have reached the maximum capacity. However, our participants had additional visual cues and a minimum of six walks through the environment, which might extend the capacity, for example, by chunking. Importantly, even if the environment was too long and contained too many targets to update at the same time in working memory parts of the environment could have already been transferred into offline long-term memory (e.g., Waller & Hodgson, 2006) and adjusted and corrected in retrospect. Adjustments to survey estimates over prolonged time have been found, for example, by Uttal and colleagues (2010) on a University campus.

Despite theoretically plausible the results of study 3 clearly show that no metric embedding took place. Observed patterns violate the metric postulate of positivity a Euclidean map is supposed to obey. According to Euclidean map theories, a place should be assigned a unique location in a coherent reference frame. Correspondingly, a place cannot possess two locations in a mental map. Thus, estimating the direction from a point A to another point B should always lead to the same pointing direction and not to such a bimodal pointing behavior as it was shown in this study. Importantly, this drifting apart of estimates for the same object is not likely to be driven by a bias towards previous estimates within a chunk of trials (i.e., target three biased towards the estimate of target two). If this would be the case, we should see similar outward biases in the possible maze group as well. However, this is not what we observe.

While the contextual-scaling model (McNamara & Diwadkar, 1997) or the category-adjustment model of spatial coding (Huttenlocher et al., 1991; Newcombe et al., 1999) were able to explain previously found asymmetries in distance estimation between pairs of places (Burroughs & Sadalla, 1979;

McNamara & Diwadkar, 1997; Sadalla, Burroughs, & Staplin, 1980) without the need to let go of the Euclidean map assumption, we reckon that they cannot be adduced to explain our results. Both approaches can explain specific asymmetric biases to the recalled location of place pairs stored in a Euclidean map when alternating the direction of the estimate (e.g., from A to B in contrast to from B to A). However, our biases occur despite constant reference-target pairing. Thus, the same context should be activated, and the same category prototype should be combined with the Euclidean metrics of the map in a Bayesian manner.

In sum, our results are in strong accordance with enriched graph theories. They postulate that metrics are stored on the local level and allow for global inconsistencies in the mental representation. This assumption is nicely reflected in the results of study 3. Even if assuming additional processes at work or a high amount of noise in the representation, still a Euclidean map representation is not able to account for the observed pointing patterns. How about considering a hierarchical representation with a Euclidean map top layer? This top layer must obey the same principles as a non-hierarchical representation, thus, acting upon the local place units and adjust their local metrics to be globally consistent.

2.4 Study 4: A hierarchy of reference frames

2.4.1 Research question

The formation of mental coordinate systems is usually detected and examined by utilizing orientation dependent memory recall (e.g., McNamara, 2003; Mou, Zhao, & McNamara, 2007; Shelton & McNamara, 2004). As described in the introduction it is assumed that a layout of objects within its surrounding space is interpreted in terms of a dominant reference direction which is then setting the direction of the mental cartesian coordinate system within which object locations are stored. Thus, evidencing a single main orientation facilitating survey estimates made from different points in space (i.e., evidencing an allocentric reference frame) can be understood as the manifestation of a Euclidean map occupying a coordinate system. Bodily or imagined alignment with the reference direction of this mental map allows for effortless retrieval of survey relations, while being misaligned requires costly transformation that finds expression in increased error and latency of the survey estimate (e.g., McNamara, Sluzenski, & Rump, 2008; Mou, McNamara, Valiquette, & Rump, 2004). Orientation-dependent survey estimates were already applied in study 1 to find a general misalignment of the reference frame for the same object layout either learned in vista and environmental space. In study 4 it was used to examine whether a hierarchy of reference frames can be formed for regionalized spaces (e.g., McNamara et al., 2008). Multiple locally confined as well as single global reference frame encompassing multiple vista spaces have been found in the literature already, indicating that a hierarchy might indeed be possible (e.g., Meilinger, Riecke, et al., 2013; Wilson et al., 2007).

The idea of a hierarchy of reference frames deviates slightly from the more rigid understanding of a single Euclidean map were all places are represented in a coherent metric format. Indeed, it is a generally accepted proposition that we cannot represent the whole world within a single cognitive map. A hierarchy of reference frames, therefore, constitutes a fair compromise that still preserves the idea of metric embedding on regional and/or global scale, while reverting to ideas of connectivity between local and regional memory nodes as proposed by (hierarchical) graph models.

In study 4 participants learned two interconnected but overall obliquely aligned regions. One region consisted of four successively connected corridors each containing a virtual object. Regions were dissociated by color (blue vs. red), semantic membership of objects (tools vs. animals), complexity of the angle of turn (90° within region, 45° at regional transition point) and spatiotemporal learning experience (longest path at regional transition point and separate learning of the regions) to trigger regionalization in memory. In a subsequent pointing task participants were teleported to different locations within the environment, always standing in the middle of a corridor. Here, similar to study 1, we manipulated the *body orientation* using one of eight body orientation in steps of 45° around the full 360° possible (e.g., looking along the corridor, looking obliquely against a wall). If local reference frames are formed participants should show a general facilitation of survey estimates when aligned with the first view experienced in each corridor compared to being aligned otherwise. If regional reference frames are formed we should find a facilitative effect when participants are aligned with a dominant region-wide main orientation (i.e., independent of whether they stand in the first, second, third or fourth corridor of a region, there should be one coherent orientation eliciting best performance). Similarly, if a global reference frame is formed there should be a single coherent main orientation across the entire environment which should enable the best performance upon bodily alignment. In short, by examining which orientations allow the fastest recall of survey estimates enabled us to detect whether local, regional and/or global reference frames are formed.

While former vista space studies already investigated in detail which factors set the reference frame orientation of a layout within a single room (e.g., Kelly & McNamara, 2008; Mou & McNamara, 2002; Rieser, 1989; Valiquette & McNamara, 2007) this aspect is clearly under-investigated in environmental space. The first perspective experienced was suggested and evidenced in a few studies (e.g., Richardson et al., 1999; Tlauka et al., 2011; Wilson et al., 2007), unfortunately the spaces used where often confounded with other factors such as parallelism of multiple corridors (i.e., simple U-shaped environments where the third corridor was parallel to the first) or differences in the duration a perspective is experienced (i.e., the first corridor was often the longest). Therefore, in order to detect potential reference frames on regional or global level we decided to test for all three factors: regional or global reference frame that aligned with the *first experienced perspective* within a region, aligned with the *most frequently experienced* perspective within a region, or aligned with the salient geometry of two *parallel legs forming a U-shape* with the corridor that is connecting both legs, and this either limited to the individual region or imposed onto the entire environment including the second learned region as well.

Besides body orientation, we manipulated the *location of the target*, which could either be in the same region as the participants current stand or located in the other region, and the *corridor distance* to the target. This allowed us, first, to see whether distance effects along local place units found in study 1 and 2 could be replicated again, and second, to expand this concept to the idea of *regional clustering*. If distinct regional memory units are stored across-region pointing should lead to longer latency compared to pointing within a region as a new memory unit has to be activated.

Just like a non-hierarchical, single layer Euclidean map representation also single layer enriched graph representations would not be able to explain a hierarchy of reference frames and the formation of local as well as regional memory units. They as well would need to be adjusted to accommodate additional levels.

2.4.2 Main results

In the analysis we concentrated on trials covering corridor distance one, two and three (maximum distance for within-region trials) to ensure comparable complexity both for within- and across-region trials (across region trials included corridor distances up to seven). First, we found main effects of corridor distance and of target region (within vs. across) on latency. With increasing corridor distance pointing latency increased and pointing took on average longer when pointing across regional boundaries. While the distance effect was also present in the absolute error, the target region had no general effect on accuracy of pointing. This indicates that—despite higher complexity and despite the fact that participants never walked across regional boundaries—the regional transition point was memorized just as accurately as the corridor angles within a region. Second[5], when pointing to targets within one's current region we found that on average participants pointed faster when bodily aligned with the first view experienced in each individual corridor compared to their performance in the remaining seven body orientations. Additionally, we found evidence for regional reference frames: Averaging performance across all corridors within a region significantly faster performance was shown when participants were aligned with the dominant reference orientation of the first corridor they had experienced within a region and when aligned with the salient parallel legs of a U-shape of a region. In contrast, when examining orientation patterns of trials involving across-region pointing neither evidence for local nor for regional reference frames could be found. Instead, now global main orientations could be identified that seemed to facilitate survey estimate upon alignment. Alignment with the perspective experienced in the first corridor of the first region learned and alignment with the U-shape geometry of the first region learned both led to best pointing performance irrespective of one's current location in the entire maze. This includes a facilitative effect of bodily alignment oblique to ones currently visible corridor when standing in the second region learned (half of the trials). It should be noted that the orientation dependent pattern for across-region trials became indefinite when including corridor distances four to seven to the analysis. Individual standard deviations suggest that this might be due to the high noise in the full across-region data.

In a second experiment, we changed the learning procedure slightly to see whether global embedding could be enhanced, and findings replicated. While in the first experiment participants always started at the outer end of each region and walked towards the transition point, now participants started learning from the transition point, walking toward the outer ends of each region (continued by back and forth navigation within a region, just as in Experiment 1). In contrast to the first experiment now it was immediately possible to relate each new corridor of the second learned region to the previously formed memory unit of the first region. Now evidence for local as well as global memory units was found. Corridor distance again had a main effect on latency

[5] We corrected for the known distance effect by using residual values from an ANOVA including only distance as a factor in order to reduce variance in the data that is already accounted for by distance.

and error and alignment with local corridors facilitated pointing latency for within-region pointings. Across-region pointings again revealed fastest estimates when aligned with the parallel geometry of the U-shape experienced in the first learned region independent on where in the entire environment the pointing was performed. However, now the picture was less clear for regional memory units. As before error was not significantly higher when pointing across regional boundaries. Additionally, this effect was also absent for pointing latency. Comparable time was needed for estimating directions to targets within ones current and in the other region. The reference frame analysis indicated a facilitative effect of alignment with a regional main orientation following the most frequently experienced perspective. However, this effect disappeared when controlling for the effect of local reference frames. Considering both analyses jointly the formation of local memory units but no regional memory units seems more likely in the second experiment.

2.4.3 Summary of study 4

We provide evidence for the formation of a hierarchical reference frame structure. Across two experiments we showed that the computation of a targets' direction is facilitated when aligned with local corridor reference directions for targets lying within one's current region, and when aligned with a single global reference direction for targets lying in the other region relative to one's current location. Local memory units seem to be accessed successively during recall irrespective of whether pointing within a region or across regional boundaries leading to the observed distance effects. Furthermore, the first experiment suggests that also intermediate levels of regions can be represented in the form of reference frames and that latency costs of activating an additional regional memory unit is to be expected when pointing across-region. These regional reference frames, however, seem to be susceptible to the learning procedure and the availability of direct reference to previously formed memory that enables linking older and new spatial information.

The absence of region effects in the second experiment contrasts other studies that were able to induce regional clustering by far less regional cues than ours. For example, semantic category of objects was sufficient to affect subsequent route decisions (e.g., Schick et al., 2015; Wiener & Mallot, 2003). It might be that different formats of regional memory (e.g., a semantic label, a topological region node connected to the other region node, a metrical embedding) can co-occur and can be targeted depending on the used task (i.e., pointing, route planning, etc.).

The results we found in study 4 are in line with a study by Greenauer and Waller (2010). They showed that two micro- and one macro-reference frame was formed for two object arrays within a single vista space. The use of the macro-reference frame was only detectable when pointing between the two layouts while within-layout pointing followed the main orientation of each individual layout (i.e., using the respective micro-reference frame). This nicely corresponds to our finding observed in environmental space. The studies demonstrate how flexible the stored memory content can be used.[6] Further, it suggests that global reference frames are only accessed when required, for example when a subordinate memory unit does not yet contain the to be recalled target location. This interpretation, however, is made with caution or instead rather be taken as a hypothesis worth testing further. If indeed no regional memory units were formed in experiment two the global reference frame should have been consulted also for within-region pointing trials according to this logic.

Further speculations can be made based on the finding that across-region pointing, on the one hand, seems to utilize a global memory unit—presumably containing relational information from all the places learned—but on the other hand, is still bound to recalling the targets in successive order corridor-per-corridor. The fact that the corridor distance effect on latency prevail irrespective of whether the target is located within or across regional boundaries suggests that the overall estimation process continues to follow the connection of local memory units along a graph structure. This aspect, along with the question of how evidencing global reference frames can be brought into accordance with findings from study 1 to 3 which suggests that no global embedding into a Euclidean map format takes place will be elaborated upon in detail in the General Discussion section.

We found that different cues have the capacity to determine the orientation of a superordinate regional or global reference frame. Not only was the perspective along the first corridor a defining factor (e.g., Richardson et al., 1999; Tlauka et al., 2011; Wilson et al., 2007), but also geometric cues (i.e., parallelism) that are not apparent at first sight but must be gathered, derived and compound across multiple corridors. This implies that both assimilation of subsequent spatial information into an initially set reference frame can be realized, but also accommodation of a new distinct reference direction. This is

[6] Observing differences in orientation dependency between survey estimates made within a region or across regional boundaries indeed suggests that multiple hierarchical levels prevail within a single participant and not only across participants (i.e., one participants forms local, another regional, yet another global memory units).

known already from vista space studies (Kelly & McNamara, 2008; Mou & McNamara, 2002; Shelton & McNamara, 2001; Valiquette & McNamara, 2007) and was suggested for environmental space as well (McNamara & Valiquette, 2004). Our study supports this and highlights that efforts should be taken to investigate potential cues setting reference frames in environmental space further.

Overall study 4 showed that memory for environmental space is hierarchical (e.g., Mallot & Basten, 2009; Stevens & Coupe, 1978; Wiener & Mallot, 2003). Results suggest that neither a non-hierarchical, single layer enriched graph representation consisting of multiple local memory units nor a single layer Euclidean map representation comprising the entire environment in a metrically embedded coherent format was formed. It supports findings showing that regional clusters are represented on a superordinate hierarchical level and affect route decisions in environmental space (Wiener & Mallot, 2003) and survey estimates when learning from figural or vista space (e.g., McNamara, 1986; Stevens & Coupe, 1978). Importantly, showing that multiple vista spaces (e.g., four from a region or eight from the entire environment) could be subsumed under a superordinate reference frame challenges the results found so far in study 1, 2 and 3, as these uniformly supported non-hierarchical, non-embedded enriched graph representations and disconfirmed Euclidean map theories. This will be elaborated upon in the next section.

3 General Discussion

In my doctoral thesis, I set out to examine the structure of survey knowledge acquired in compartmentalized, walkable environmental space. I contrasted two major approaches and their potential combinations that are discussed in the literature: The formation of an allocentric, globally consistent Euclidean map that assigns coordinate values from a mental coordinate system to each place visited (e.g., Gallistel, 1990; O'Keefe & Nadel, 1978; Siegel & White, 1975; Thorndyke & Hayes-Roth, 1982), and the representation of an enriched graph preserving the local places and their connectivity as individual units, and augmenting this local place-to-place information with local metrics of angles to turn and distance to walk to get to the next place (e.g., Chrastil & Warren, 2014; Meilinger, 2008; Warren et al., 2017). I derived several new predictions from both approaches that I explored in four studies to illuminate the debate.

3.1 Main findings

In sum, I could show the following:

(1) Vista and environmental survey knowledge are different, and these differences have their origin probably in the presence of opaque barriers (and not so much in the successive encounter of spatial information or the travel along a predefined path) structuring the environmental space and thereby compartmentalizing the representation (study 1). This indicates that vista spaces serve as distinct units in environmental space survey knowledge.

(2) These units are likely to be composed of local reference frames that are representing the spatial properties within a single corridor or room in a uniform format, as was indicated by orientation dependent memory recall following the local corridor geometry (study 1 and 4).

(3) Survey estimates seem to be constructed on the fly when needed, involving a time-consuming successive activation of individual, local memory units. This was indicated by the place-to-place distance effect on latency (study 1, 2, and 4). The more local place units are activated, the more time it takes to come up with an estimate of a target direction. The distance effects and the order of layout reconstruction (study 1) further suggest that this successive estimation process is not driven by the Euclidean straight-line distances between places but is bound to the connectivity of the graph which in turn is based on the egocentrically experienced place sequence during learning.

(4) The egocentric learning experience does not only seem to determine which memory units are connected by links but further seems to specify the local translation and rotation information in a directed fashion (i.e., precise information how to get from A to B but not necessarily how to get from B to A). As suggested by the route direction effect on survey estimates (study 2), this facilitates the speed of processing along the specified direction of the link that can be driven, for example, by the direction of a learned route through the environmental space.

(5) Survey estimates are based on place-to-place metrics which are not necessarily metrically embedded on a global scale. Instead, each local place-to-place metric can be subject to very individual distortions. As indicated by the impossible pointing behavior of study 3, deviating estimates can occur depending on which mental route along a graph one takes. This clearly violates Euclidean metric postulates.

(6) Survey estimates are transient and generally need to be constructed again when queried again later in another trial (replicating corridor distance effects across the entire testing phase) but are available at least for a short amount of time to base contiguous consecutive pointings on. This was reflected in shorter and relatively stable latencies for trials succeeding the initial estimate within a chunk of related, neighboring targets (study 2 and 3). In those cases, potentially only the difference vector from the previous estimate to the new target had to be computed, thus, involving the additional activation of only a single new memory unit.

So far, all these observations are very informative about the sub-layer of survey representations, specifying neighbor-to-neighbor place information. Additionally, I also detected manifestations of vista space clusters and facilitated pointing performance when aligned with orientations on the regional and global scale:

(7) Survey knowledge of environmental space seems to be subject to clustering and global consolidation. This was reflected, for example, in negative effects on latency when another regional unit was activated (region-to-region effect) and orientation dependent recall on regional and global scale (study 4). This indicates that consolidation into a superordinate memory unit occurred which covers more than just a single vista space. Access to this additional information seems to be de-

pendent on the position of the target relative to the navigator, as pointed out by the finding that global main orientations only seemed to be used when pointing across regional boundaries (study 4).

Before continuing with the theoretical embedding I shortly want to summarize how my data corresponds to previous findings in the spatial cognition literature. My results strongly highlight the particular role of a single enclosed space for structuring our spatial survey memory. They extend studies showing, for example, effects of opaque barriers on distance estimations (e.g., Cohen, Baldwin, & Sherman, 1978; Kosslyn, Pick, & Fariello, 1974; Newcombe & Liben, 1982) and disturbed updating of locations beyond one's current local environment (e.g., Wang & Brockmole, 2003b, 2003a), and nicely corresponds to previous findings of local, corridor-bound reference frames (e.g., Meilinger, Riecke, et al., 2013). Evidencing route forward facilitated pointing performance shows that the route direction effect that is usually associated with route knowledge (e.g., Janzen, 2006; Schweizer et al., 1998) can also be generalized onto survey knowledge. This indicates that previous studies showing uncorrelated error for forward and backward pointing (e.g., Meilinger et al., 2018) might not reflect the formation of two uniquely distorted mental maps but one graph format that is directed. The impossibility of pointing patterns in the impossible maze group of study 3 is in accordance with but also extends previous findings that seem to violate the Euclidean metric postulates maps must obey (e.g., R. W. Byrne, 1979; Moar & Bower, 1983; Tversky, 1981)(Burroughs & Sadalla, 1979; R. W. Byrne, 1979; McNamara & Diwadkar, 1997; Moar & Bower, 1983; Sadalla et al., 1980; Tversky, 1981). In contrast to these previous findings, we are confident that our results cannot be accounted for by category knowledge (e.g., Huttenlocher et al., 1991) or by landmark saliency (McNamara & Diwadkar, 1997), but indeed reflect non-embedded survey memory. The hierarchical structure of memory with levels that can be flexibly accessed and operated on has been observed for direction estimates made within and between multiple layouts learned in a single vista space (e.g., Greenauer & Waller, 2010). We show that similar processes and memory structures are employed for clustered multi-corridor spaces, thereby supporting previous findings and models postulating hierarchies in spatial memory for environmental space (e.g., Stevens & Coupe, 1978; Wiener & Mallot, 2003).

3.2 The structure of survey knowledge

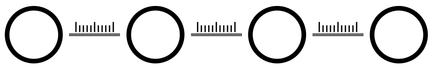
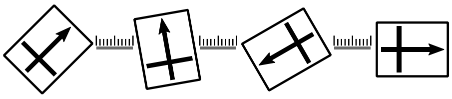
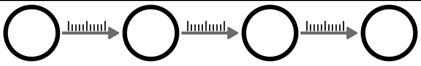
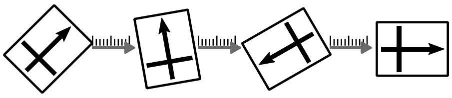
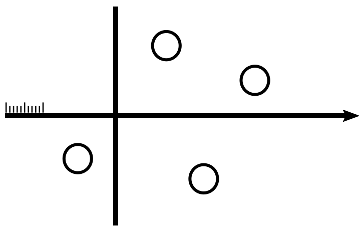
What conclusions can we draw based on the effects summarized above with regards to the possible representational structures of environmental space survey knowledge? In Table 2 and 3 I summarized the observed effects again and contrasted them with a number of feasible survey memory structures that are based on the theoretical concepts presented in the introduction. The models that are visualized in the tables represent memory structures for four places learned in environmental space (e.g., four corridors or streets traveled successively). Check and cross marks in each open cell indicate whether the respective model is able to account for the effects found in the studies of my thesis. Violations (cross marks) are additionally colored in red. I would first like to concentrate on the local place level of representation (Table 2) and describe and discuss the more complex models that consider additional information about regions and potential global embedding on superordinate levels (Table 3) further below.

3.2.1 Representation of the place-level

Model #1 to #5 in Table 2 concentrate on the description of a single hierarchical layer of the representation, the sub-layer. Here I compared enriched graph representations with a Euclidean map memory. More precisely, for the enriched graph representations I contrasted feasible models that assume a local place to be represented in a local reference frame format (model #2 and #4) (e.g., network of reference frame proposed by Meilinger, 2008) and those that do not expect a local main orientation on the place level (model #1 and #3) (e.g., local graph model by Chrastil & Warren, 2014). Further, I added the assumption of non-directed (model #1 and #2) or directed graph edges/links (model #3 and #4). The Euclidean map representation consists of a single coordinate system specifying memorized locations as coordinate values (e.g., Gallistel, 1990; O'Keefe & Nadel, 1978).

In sum, the non-hierarchical Euclidean map representation (model #5) is not able to explain the observed local effects of my studies. Euclidean map theories typically postulate that long-term survey memory stores places relative to a common coordinate system. All information should be readily accessible in this single representational unit and thus enable an all-at-once readout of the spatial information or at least an immediate computation thereof. Such a format by itself would not predict a successive activation of local place units along a learned order. Euclidean map approaches also don't specify whether a location in the mental map can possess its very own local reference frame. There is also no need for successive estimates to be based on one another, as calculating the direction between my current location and the new target should be just as fast as (or even faster than) using the old estimate and add another vector information towards the next target to it.

Table 2. How is the place level structured? Contrasting the local effects found in the four studies with the different theoretical approaches.

#	Place level	Successive activation of memory units evoking place-to-place effects	Orientation dependency on local place level	Facilitation of estimates along route direction	Estimates along learning order	Progressive estimates during successive pointings
Enriched graph models						
1		✓	×	×	✓	✓
2		✓	✓	×	✓	✓
3		✓	×	✓	✓	✓
4		✓	✓	✓	✓	✓
Euclidean map model						
5		×	×	×	×	×

Note: The scale attached to the node links in model 1-4 represents the local place-to-place metrics. In model 5 globally consistent metrics are expected. ✓ The model can explain the observed effect. × The model cannot explain the observed effect.

In contrast, all these observations can be explained by assuming a single layer of an enriched graph and a successive usage of spatial information along that graph structure to construct survey estimates to distant objects. Further, our results allow specifying that the local units that are stored each possess their own locally confined reference frame and that the links between places specify their connections in a directed fashion[7]. Hence, considering all these

[7] In contrast to the Euclidean map approach, a non-hierarchical enriched graph representation could also account for the impossible, non-Euclidean pointing behavior found in study 3. While the former requires globally coherent embedding, the latter does not. However, I decided to discuss

effects the Euclidean map (model #5) and the enriched graph representation models #1, #2 and #3 can be ruled out, rendering model #4 the best theoretical account for the place sub-level of survey memory based on our results (marked in green). Such a directed graph was proposed by Meilinger (2008). However, such a non-hierarchical representation cannot explain effects of regionalization and consolidation on the global scale. Hence, it seems necessary to assume that multiple local places can be subsumed in a joint memory unit.

3.2.2 Representation of non-local information

In the next step, I would like to consider the effects that are pointing towards additional layers of information, which might, for example, be represented on a superordinate level in a hierarchical memory structure, and evaluate which representational structure could account for them. Table 3 summarizes the models discussed in the following. Due to the lack of space the visualization of the different layers only includes two levels, yet many more are conceivable (e.g., local, regional, global).

First, I want to discuss the potential interaction of hierarchical memory layers. If a target lies outside of a formed memory unit, for example, outside of the place unit or region unit that is stored, and at the same time a higher order memory unit, for example, a global memory unit is accessible, will the entire recall process be entirely based on this superordinate level thereby ignoring subordinate level(s)? I reckon this is not the case. Study 4 shows that albeit evidencing an orientation dependent recall following a global main orientation (i.e., potentially activation of global memory unit), coming up with a survey estimate is still bound to corridor distance effects, hence, following the local place-to-place connectivity probably stored on a subordinate level. Hence, even if higher order memory units are formed and accessed during recall, the subordinate place-level is still co-creating the survey estimate. Correspondingly, any hierarchical theory we suggest and discuss now should be bound to this restriction, the simultaneous use of subordinate place levels if superordinate levels are called upon.

Model #6 in Table 3 proposes a hierarchical representation consisting of a purely topological, subordinate place level and a Euclidean map as a superordinate level, hence, all metric information is stored on the higher level, while the subordinate specifies connectivity between places only. This means that no metric information about translation and rotation are specified on the local

---

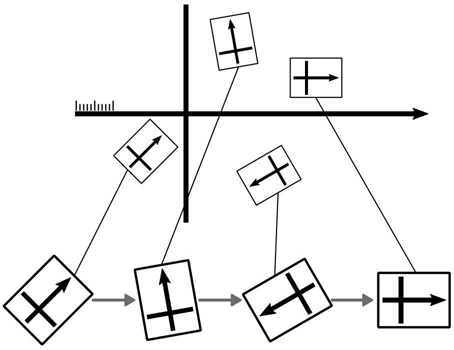
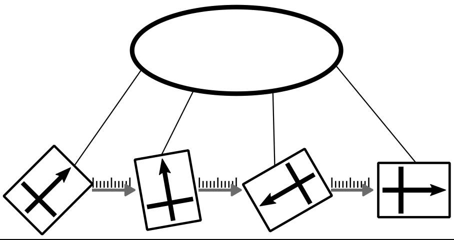
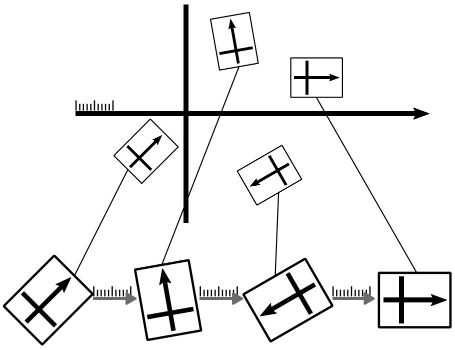
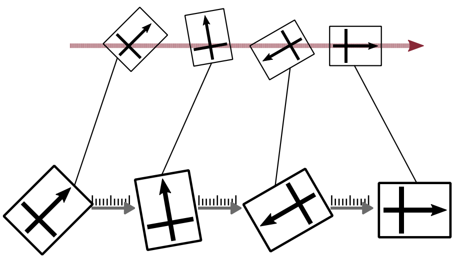
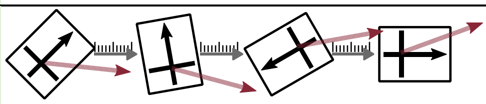
the impossible pointing behavior in Table 3, as I reckon it to be of major importance to differentiate the possible multi-level models.

level of this model. Such a representation could indeed account for place-to-place distance effects and a successive recall (which both have been discussed in Table 2) if assuming as well that retrieval of survey relations is bound to activation along the topological connectivity on the place level. If the underlying route topology includes directed links direction effects are feasible as well. However, such a representation implies that properties of the external space are preserved in a metric format that complies with Euclidean metric postulates on a global scale. Yet, the distorted “impossible” direction estimates observed in study 3 violate the metric postulates of Euclidean maps. Hence, model #6 could not account for the bimodal pointing patterns found in study 3.

Following the deduction from the last two paragraphs two points seem to be vital when trying to come up with a potential representational format that can account for our results: First, the local place level is always involved in survey estimation processes even if superordinate memory units are consulted. Second, there must be metric information stored on this local place-to-place level. Hence, the lowest place-to-place level in the remaining models #7 to #10 (also Table 2) corresponds to the deductions expounded in the last paragraph: a network of local reference frames connected by directed links enriched with local metrics.

Model #7 is a graph model with a topological superordinate level loosely based on the hierarchical model that was proposed by Wiener and Mallot (2003). All metric information is stored on the local place level, but clustering of individual places can be achieved by forming an additional layer of region nodes, and potentially also a global node covering all places. The connectivity of the superordinate nodes is represented purely topological. Each place, besides being connected to its neighboring place, is also connected to its corresponding regional node on the superordinate hierarchical level. Such a representation is able to account for the increase of latency when pointing beyond regional boundaries when assuming that the superordinate regional memory units are co-activated and incorporated in the estimation process (study 4). Furthermore, since the superordinate regional nodes reflect topological information about connectivity to neighboring regions but do not specify metric information, all metrics are stored on the local place level and are not required to be globally embedded. Hence, this model would also be able to capture the impossible pointing behavior of study 3. However, as the superordinate memory unit is not specified further a pure topological node of

Table 3. What is represented beyond local place-to-place information? Contrasting the remaining local, regional and global effects found in the four studies with the different theoretical approaches.

#	Beyond local place nodes	Distorted, globally 'impossible' estimates following local metrics	Successive activation of memory units evoking region-to-region effects	Orientation dependency on regional and global level
	Euclidean map without local metrics			
6		×	✓	✓
	Topological node clustering place units			
7		✓	✓	(×)
	Hierarchy of reference frames			
8		×	✓	✓
	General reference direction			
9		✓	✓	✓
10		✓	✓	✓

Note: The scale attached to the node links in model 7-10 represents the local place-to-place metrics. No local metrics are expected in model 6. In model 6 and 8 globally consistent metrics are expected. ✓ The model can explain the observed effect. × The model cannot explain the observed effect.

a region could not account for the orientation dependent recall on the regional and global level. It should be noted here that although this model cannot account for the entirety of my findings I do not mean to imply that regions or higher order memory units can only be represented in an orientation-dependent format. As pointed out in the summary of study 4 already (aiming to explain the discrepancy between the vanishing region effect in Experiment 2 and previous studies evidencing the formation of regions with far less salient cues than in my study) different formats of regional memory units could co-occur (e.g., a semantic label, a topological region node connected to the other region node, an orientation-dependent format) and triggered by different tasks.

Model #8 is a hierarchy of reference frames. Such a model would be a compelling hybrid version of enriched graphs and Euclidean maps as it assumes multiple Euclidean maps that differ in the space they encompass, and it includes the global embedding of multiple locally confined places. Assuming that both regionally confined Euclidean maps for a subset of places were formed on the regional level of the hierarchy as well as a global, all-encompassing Euclidean map on the superordinate level, this model is able to explain the orientation dependent recall following regional as well as global main orientations (study 4). It can also account for the latency increase for estimates across regional boundaries since another regional memory unit must be activated in a time-consuming manner. This, at first sight, renders the hierarchy of reference frames a compelling model. It indicates that the subordinate memory units indeed have the characteristic of a Euclidean map with a main axis that facilitates estimates upon bodily alignment. This interpretation, however, contradicts strongly with the finding of "impossible", deviating direction estimates following non-embedded local metrics (study 3)[8]. So, given that the same learning mechanisms underly the representation of the environments of all my studies and the same representational structure is aimed for in the spatial memory system, how can a Euclidean coordinate system be formed (as indicated by study 4) without globally embedding the learned locations (as indicated by study 3)?

3.2.3 A general global reference direction

To bring together both findings, I reckon that it is important to first reflect on what we measure in orientation dependent recall scenarios. Deduced from the

[8] Not only does this last finding contradict the formation of a Euclidean map that is possessing an oriented mental reference system, but it also contradicts models assuming an orientation-free global embedding (e.g., Sholl, 2001) because both approaches require a representation which is globally consistent.

Euclidean map approach places should be stored in a mental coordinate system with set cartesian axes that orient relative to dominant spatial cues in the environment. This, in turn, should manifest in main orientations facilitating survey estimates that are detectable in orientation dependent recall scenarios (e.g., Shelton & McNamara, 2001). Deduced the other way around, evidencing orientation dependence in survey estimates may not always necessarily involve an embedded, coherent coordinate system where all spatial relations are explicitly specified. Alternatively, the "conceptual north" that was measured in study 4 might reflect an anchor orientation propagated across multiple corridors, for example, via a global sense of direction system (e.g., Sholl, Kenny, & DellaPorta, 2006). For each local memory unit participants might have had access to an additional vector denoting a global direction which might be stored on a superordinate hierarchical level. Such a *general reference direction* in additional to local charts has been proposed already by Poucet (1993), who wrote "the distinct reference directions provided by different local charts [may] be combined into a single, local chart-independent, overall direction so that a two-stage vector summation would be sufficient for correct orientation to a distant environment" (p. 173). Both Poucet (1993) and the results from study 4 (local and regional orientations facilitate memory recall for within region pointing, global orientations facilitate survey estimates across regional boundaries) implicate that this general reference direction is represented in addition to the local reference frame. Such superordinate vectors can facilitate the coordination and alignment of the local memory units stored on the subordinate level and therefore help during the successive construction of a survey estimate across the graph. Model #9 and #10 visualize two feasible versions of such a memory structure (general reference direction depicted in red), one being explicitly hierarchical (#9) the other one not necessarily, as the general reference direction is attached to each local memory unit (#10).

Let's consider both in detail. Are they sufficient to account for orientation dependency on regional and global scale (study 4) but at the same time allow for a globally inconsistent representation (study 3)? Model #9 stores the general reference direction on the superordinate level of the hierarchy. Each place of the local level is connected and explicitly aligned with respect to this one general direction. This implies that—just as for the hierarchy of reference frames that are aligning local reference frames with a superordinate main direction—this alignment with the general reference direction vector should propagate back to the local level and correct and adjust the defective local place-to-place metrics that are stored in long-term memory. What would this imply for the circular impossible maze in study 3? Figure 2 gives an example, standing at the shoe and walking from there to key and duck. As the left side of Figure 2 shows, depending on whether one decides to walk there clockwise or counterclockwise following the impossible local place-to-place metrics one ends up at different locations in physical space and the local alignment of the corridor also differs relative to the physical world outside of the virtual reality. This is the same for any location one wants to walk to in the impossible maze. According to model #9, an alignment of the seven corridors should be achieved and specified relative to a reference orientation on the superordinate level. The red arrow attached to each local reference frame in Figure 2 left visualizes the perfect, physically correct global main orientation a participant could try to propagate over the entire environment during walking when trying to keep oriented relative to a mental "north". It becomes clear that this perfect general reference direction is impossible when faced with this impossible maze. Take, for example, the clockwise and counterclockwise orientation of the general reference direction relative to the local reference frames of duck and key. They point to different directions relative to the local corridors. In order to follow coherently a general reference direction, this conflict must first be reduced by trying to match the alignment of the corridor pairings in memory. An example is given in Figure 2 right again for duck and key corridor (remaining corridors adjusted as well but colored in grey as their partner corridor for the cw and ccw direction is not visible in this figure). It shows that the two possible alignments of the partner corridors relative to a global direction should be matched, for example, by averaging the orientation of the local reference frames relative to the general reference direction. For the remaining corridors, the same alignment of each of the two impossible corridor orientations should be tried to achieve as well. This attempt to align all seven corridors with respect to a single global reference direction should then propagates from the superordinate to the subordinate local level and requires the adjustments of all *angular* place-to-place metrics (i.e., how much I must rotate my body from here to the next neighbor). Basically, trying to match a general reference direction to this circular layout of seven places should lead to a sum of inner angles between places of 900°, which is that of a heptagon. However, only angular metrics should be affected by this global alignment. The translational distance information, however, is not specified and also not directly affected by the general reference direction, just as visualized in Figure 2 right. This implies that, although the orientations of the local units are fitted, the resulting local place-to-place metrics (translation and rotation to the next neighbor) can still be globally incoherent and the pointing pattern can reveal that "impossible" locations are stored.

Model #10 holds that the general reference direction is stored directly on the local level, attached to each memory unit. This model is very similar to model #9 as it also requires a coherent general reference direction relative to which each local reference frame is defined (similar to Figure 2 right). Importantly and in contrast to model #9, due to the local confinement of the general reference vector there is no need to adjust the local place-to-place metrics stored in long-term memory (i.e., how I must turn to be aligned with the neighboring corridor geometry) to fit the rotational information that can be computed from the rotational deviation of neighboring vista spaces from their attached general reference direction. Hence, just as local place-to-place metrics can be globally inconsistent the same holds for the locally stored general reference direction. In other words, the model implies that there can be nonconformity between the rotation information stored as part of the local place-to-place metrics and the rotational information implicitly carried by the orientation of a corridor and its neighbor relative to their attached general reference direction. The general reference direction is supposed to aid the coordination and alignment of the local memory units during recall. Hence, here it is likely that the general reference direction interacts with the defective local metrics during the estimation process and adjusts (but not overwrites!) the estimated direction to a target, thereby potentially bringing it closer to a globally possible picture. But like model #9 global embedding is not achieved.

Taken together, the difference between the models is, that #9 adjusts the rotational metrics of the local place-to-place metrics that are stored on the subordinate level in long-term memory to comply with the superordinate reference direction, whereas #10 retains the local metrics as experienced but they must be integrated with the additional general reference frame information during the estimation process, hence, during memory retrieval. Both models could explain the "impossible" pointing patterns found in study 3 and combine these with the orientation dependent recall following regional and global orientations found in study 4. One could argue that the hierarchical #9 sets more limits to the "impossibility" that can be represented as local rotation must be brought in coherence already in long-term memory, while the non-hierarchical #10 allows for storing larger discrepancies without necessarily inducing conflict and adjustments be done on the representation itself. Both approaches could explain why pointing patterns in the impossible maze group of study 4 were very close to the impossible local metrics prediction, but not perfectly match it. The use of a general reference direction could have helped to adjust the local metrics during pointing to match with the fact that the space was circular and that the loop should close at some point. Hence, the general reference direction helped to create a pointing pattern that approaches the possible case but without global embedding or the obedience to metric postulates.

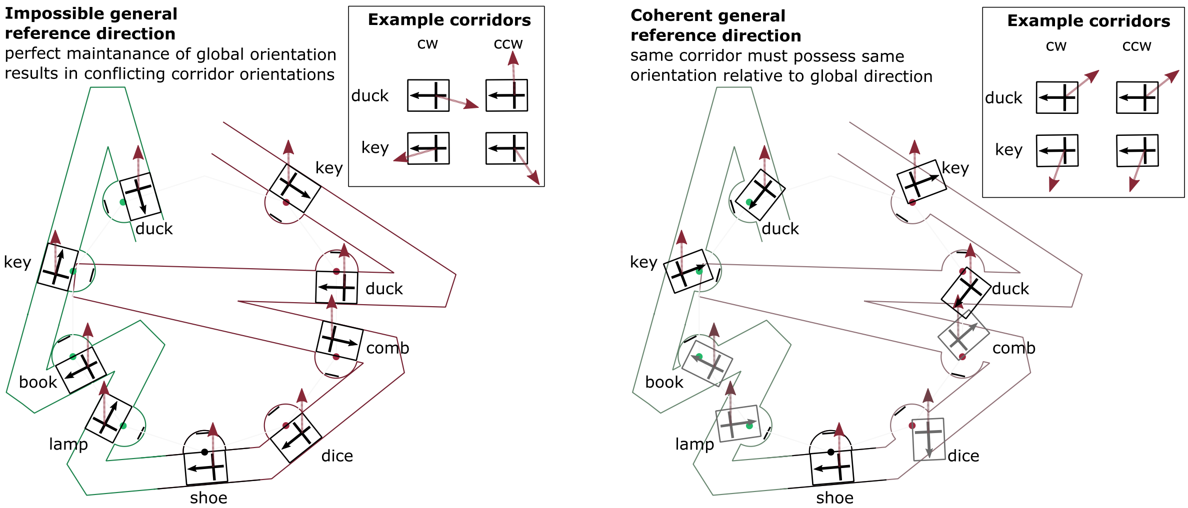


Figure 2. How could the two models proposing a general reference direction handle the layout of the circular impossible environment used in study 3? The layout shows that depending on whether a navigator travels clockwise (cw) or counterclockwise (ccw), for example, from shoe to key/duck, they end up at different locations in real physical space. Also, the orientation of the clockwise-counterclockwise corridor pairs differs. **Left:** Propagating a mental "north" across the impossible environment (red arrows) leads to an impossible general reference direction for each corridor. The example corridors show the difference between the orientation of the reference direction relative to the local reference frames for the same corridor either walked cw or ccw. **Right:** This conflict must be reduced by generating a coherent general reference direction. This can be done, for example, by averaging the orientation of the local reference frames from both corridors relative to the general reference vector (red arrows). The example corridors show how the vectors are now matched. Deduced from model #8 in Table 3, if a general reference direction is used as a superordinate memory layer which aligns the subordinate local reference frames (i.e., all corridors), this should lead to adjustments of the rotational components of all place-to-place metric in long-term memory, as visualized by the rotated local reference frames relative to the actual corridor orientation. The actual position of the local unit, however, is not directly affected and allows for global inconsistency in pointing behavior. Grey reference frames are adjusted as well but their partner corridor for the cw and ccw direction is not visible in this figure. Model #8 holds that this matching process is only affecting the recall process but not alter long-term memory.

In sum, such a general reference direction approach can account for the entirety of findings obtained during my doctoral project and described in this thesis (hence, marked in green in Table 3). Is it possible to determine based on study 4 alone whether one or multiple general reference direction can be added to a local memory unit? The fact that regional orientation dependency was found for within region pointing and global orientation dependency for across region pointing suggests that these patterns are no reflections of inter-individual difference but instead that multiple general reference directions can be stored and used flexibly when needed. An experiment allowing for detailed analysis of pointing patterns of a single individual would be necessary to verify that multi-level information indeed exist within-subject, or a larger sample size that allows for the detection of potential subgroups (e.g., a group storing one global reference direction vs. a group storing general reference directions for a region of corridors). Importantly, the orientation dependency on regional and global level and the non-embedded, impossible local metrics (and of course the other effects described before) have been observed in separate studies of my thesis. To experimentally verify my conclusions that a general reference direction is stored I reckon it a fruitful and necessary endeavor to combine orientation dependent testing and impossible worlds in a joint experiment. Showing a main orientation facilitating survey estimates in a non-embedded, globally impossible environment would be a convincing argument for the general reference direction model.

Even though my joint results question the formation of mental coordinate systems covering and embedding multiple vista spaces they do not oppose the common view that mental coordinate system can be formed within vista spaces. Much research has shown the formation of a single (e.g., Shelton & McNamara, 2001) or even multiple reference frames (e.g., Greenauer & Waller, 2010; Mou, McNamara, & Zhang, 2013) for object arrangements within vista space. Especially, also study 1 of the current thesis implicates that these reflect metrically embedded representations. In contrast to the environmental space group the vista space group in study 1 accessed all objects of the layout with similar ease from memory for making survey estimates. Latency of pointing was not depending on Euclidean straight-line distance between participants current position and the target. This favors a simple process of reading-out coordinates from an embedded representation and calculating a difference vector to the target when learning takes place in a single vista space.

3.2.4 Is global embedding possible?

In the light of the reviewed literature and appreciating the results I obtained and discussed in my thesis it seems valid to argue that the survey knowledge acquired in environmental space is not represented in the format of a Euclidean map in its literal and rigid interpretation of a global embedding of all places encountered in a common frame of reference. Nevertheless, it should be added that based on the four studies I presented here I cannot exclude the possibility that global embedding and the formation of a Euclidean map—at least as a superordinate memory unit in addition to a local graph—might take place under certain conditions.

Could extensive exposure help to form a Euclidean map? We know from rodent grid cell recordings that firing patterns that are initially disrupted can form a continuous representation spanning the entire compartmentalized box after a prolonged experience of ca. 20 days (Carpenter et al., 2015). In contrast, Ishikawa and Montello (2005) showed that learning two interconnected routes in a city neighborhood over 10 weekly sessions resulted in a slight continuous improvement in spatial judgments for some participants, but also large interindividual differences. While some participants had acquired accurate survey knowledge from the beginning others retained a weak representation throughout. Even though Ishikawa and Montello did not specifically test for global embedding their findings show that more exposure to an environment does not necessarily improve the representation. It remains unclear what time can do for forming a Euclidean map. A possible and compelling experiment could involve having people repeatedly learn an impossible environment and testing one group of participants after the first session, another group of participants only after a week or a month full of learning sessions. Comparing their rate of global embedding compared to the bias towards impossible local metrics could give insights about the time course of the potential formation of Euclidean maps.

Could global embedding be achieved with less complex environments? The most compelling case against a global embedding on any level of the representation was made in study 3. This was also the most complex environment with non-orthogonal angles and corridors of different length. However, indications of impossible, non-embedded mental representations were also observed in much simpler impossible environments, for example in the impossible triangle and rectangle used by Kluss and colleagues (2015). Based on these results in combination with the corridor distance effects found in study 1, 2 and 4 (all of them used simpler environments than study 3) it seems that even when confronted with very simple and small environments the representation remains to be compartmentalized and is probably not brought into global coherence.

Could global landmarks facilitate global embedding? The aim of my doctoral project was to focus on the representation of environmental space under unaided conditions, forcing participants to bring together spatial information from clearly circumscribed, mostly encapsulated local environments. Therefore, I chose to prevent access to any global landmarks visible from multiple local places in my experiments as these could have aided global embedding. Head direction cells in rats, for example, are sensitive to the location of distant landmarks surrounding their environment (e.g., Taube, Muller, & Ranck, 1990; Winter & Taube, 2014). This suggests that global landmarks can be utilized as a general reference direction. More precisely, having access to a stable global landmark even makes it unnecessary to stay globally oriented, for example, based on a self-created mental general reference direction that I was describing above. Although human spatial cognition literature was not able to show a general advantage in having access to global landmarks compared to local landmarks for correct route decisions (Steck & Mallot, 2000) and survey estimates (Meilinger, Schulte-Pelkum, Frankenstein, Berger, & Bülthoff, 2015), there are findings indicating that the reliance on global landmarks allows for more flexible wayfinding behavior in cluttered environments (Hurlebaus, Basten, Mallot, & Wiener, 2008). However, the effect of global landmarks on the structure of survey knowledge (i.e., global embedding) was not yet examined. Correspondingly, learning an environmental space from looking at a map is different from learning from egocentric navigation experience only. In case of map learning indeed a Euclidean map representation can easily emerge based on the mere fact that the spatial input itself was already presented in this exact format.

Despite these potential factors that might hinder or facilitate global embedding what clearly remains from the four studies I presented here is, that in order to make survey estimates in environmental space a globally consistent Euclidean map is not required. We seem to be prone to store spatial information gathered in environmental space in a piece-wise fashion and we can base our survey estimates on this knowledge by incrementally recalling all the necessary information on the fly.

3.3 Neuronal correlates of mentally walking a Euclidean map?

The neural structures that have been interpreted to be the brains spatial administration center and to allow Euclidean metric maps to form all group around the hippocampus. Especially grid, place and head direction cells are assumed to be the neural underpinning for the formation of a metric Euclidean map (e.g., Gallistel & Cramer, 1996; McNaughton et al., 2006; O'Keefe & Nadel, 1978). While place cells identify unique locations in space head direction and grid cells provide the essential rotational and translational metrics to store. The contingency of a Euclidean map covering an entire environmental space is supported for example by the finding that grid cells in rodents previously dissected in their activity patterns across multiple compartments can realign and form a uniform pattern after sufficient exposure, similar to the pattern observed in vista spaces (Carpenter et al., 2015). It has been proposed that the same neural structures that support the construction of a mental representation of space are used for memory consolidation but also for upcoming navigation including route planning to a target not currently in sight (e.g., P. Byrne, Becker, & Burgess, 2007b; Sanders, Rennó-Costa, Idiart, & Lisman, 2015). Thus, a *mental walk* through the previously learned environment is simulated or imagined so as to explore potential routes or extract goal-directed heading vectors. Electrophysiological measures in rodents detected a candidate mechanism for such a mental walk (e.g., Bush, Barry, Manson, Correspondence, & Burgess, 2015; Erdem & Hasselmo, 2012) which is often referred to as *preplay*. During rest periods of the rat a subset of place cells fire in a fast succession reflecting upcoming trajectories to be traveled (e.g., Dragoi & Tonegawa, 2011). This prospective activity during preplay of place cells was further found to be correlated with coherent grid cell activity (Ólafsdóttir, Carpenter, & Barry, 2016). Interestingly, such replay is not limited to recalling the precise path that has been travelled before but can also be observed covering novel, previously untraveled paths, which for example combine parts of an environment experienced separately, or involve areas that have been only visually explored but never travelled before (e.g., Ólafsdóttir, Barry, Saleem, Hassabis, & Spiers, 2015; Gupta, van der Meer, Touretzky, & Redish, 2010). This indicates that not just simple stimulus-response strategies are at play but instead a flexible utilization of a network of spatial information, which admits the resolution of survey tasks.

Such mental simulations also seem to occur in humans. In fMRI studies a 60° directional periodicity of BOLD-signal modulations in the entorhinal cortex interpreted to reflect grid-like signals, has been found during navigation of a virtual environment but also during imagined movement through that space (Horner et al., 2016). Likewise, neural sensitivity to one's head direction during mental imagination of different viewpoints was observed (e.g., Bellmund, Deuker, Navarro Schröder, & Doeller, 2016). On top of that, neural activity patterns where observed that correspond to the direction of a target relative to one's current location and one's viewpoint (e.g., Chadwick, Jolly, Amos, Hassabis, & Spiers, 2015) as well as activity patterns that seem to reflect not only route distance but also straight-line distance to the target (e.g., Howard et al., 2014). Such findings can be interpreted as evidence for a Euclidean map directly representing object-to-object metrics across the entire environment in memory (e.g., Epstein, Patai, Julian, & Spiers, 2017). However, an important counterargument can be brought forward. The BOLD patterns observed in these studies were based either on memory for vista space, environmental space learned via maps (i.e., direct access to a Euclidean map from a bird's eye perspective) or environmental space with central and surrounding global landmarks (i.e., allowing to maintain oriented with respect to viewing orientation). Hence, these studies might mirror the human brain's general capabilities under certain inputs, but importantly generalizability of these Euclidean map references to environmental space memory that is acquired based on our natural spatial capabilities and without distinct global cues is doubtful.

Interestingly, although advocating the representation of a Euclidean map in memory the mental walk mechanism proposed to explain the recall thereof is following the succession of egocentric experiences along traveled paths instead of the read-out of coordinates or directly represented distance and angular information from a completed, all-encompassing mental map. Not only the mental walk approach but also the replay patterns observed in rodents suggest that the recall of the paths to the goal is following physically traversable paths through the environment. To my knowledge to date, no replay patterns could be observed that followed a straight line towards a goal thereby "passing over" opaque walls—which indeed would make a nice case for a hard-wired straight-line relation between distant places.

The mental walk approach relates well to the mental processes assumed by enriched graph representations. Meilinger (2008), for example, suggested a construction of a mental model involving the successive activation of the non-visible parts of the environment vista-per-vista space from one's current location along the imagined path to the goal to come up with a survey estimate. Both the mental model from Meilinger (2008) and the mental walk approach (e.g., Byrne, Becker, Burgess, 2007) could account for the latency increase with increasing corridor/place distance observed in study 1, 2 and 4 of the current thesis. However, whereas Meilinger's theory assumes an increase with every new vista space memory unit activated, the mental walk approach indicates that latency should increase with every meter that was mentally traveled. A prediction worthwhile testing[9] and an issue worthwhile discussing in more detail in the next section

[9] See O'Malley, Bülthoff and Meilinger (2014) who presented an experiment aiming to differentiate both approaches. Their findings suggest that both the number of local vista spaces but also the travelling distance along these vista space seem to affect performance in survey estimates.

Despite the commonalities between the mental walk and the mental model approach, they differ on a major assumption. In contrast to Meilinger (2008) who's model particularly opposes the global embedding of places in environmental space into a Euclidean map, advocates of the mental walk usually assume that this process is based on recalling spatial information from a Euclidean mental map (e.g., Byrne, Becker, Burgess, 2007; Nadel, 2013). As explained in detail in the last section the results of my studies clearly violate Euclidean map approaches. Hence, based on my studies I reckon there is a need to relax the assumption of a Euclidean map but instead face the possibility that a mental walk might be done based on a non-Euclidean representation.

3.4 The definition of a place

Related to the distinction between mental walk and mental model is the question of what we actually define as a "place". Indeed, the aspect of what exactly we store as place-memory units should be discussed when advertising the existence of enriched graph representations. As Mallot and Basten (2003) pointed out nicely "The notion of a 'place', unanimous as it may seem, is not easily defined" (p.1662). In my doctoral thesis in four studies, I showed the importance of local memory units. For example, they possess their own reference frames and seem to be activated in succession in a time-consuming manner. However, the "local units" were not specified consistently across all studies. In study 1, 3 and 4 we decided to designate individual corridors (i.e., vista spaces) containing one or two objects as our local memory units. In these studies, standing in the middle of each corridor did not provide any visual information about the neighboring corridors and the objects within except for the visibility of the turning angle. Also, no route decisions had to be made during learning as the corridors basically were constructed in a long zig-zagging pipe without intersecting corridors. In contrast, study 2 used a succession of intersections (each requiring a route decision to make) as "local units" for the analysis of place-to-place distance effects and route direction effects. Here neighboring intersections were visible from one's current view when teleported to an intersection. Both versions of places produced place distance effects.

So, what in our stimulus-rich environments induces the processing of place information in the first place? As described above, in the neuroscientific literature a place is defined by the firing patterns of a place cell, which declines gradually with increasing distance from its ascribed place, thereby overlapping with the firing fields of "neighboring" place cells (which are not necessarily located side-a-side in the brain). Thus, place fields cover the entire environment in a continuous manner. A single place cell in the CA1 region of the hippocampus in rodents can cover an area from 20 x 20 cm (dorsal part of hippocampus) via 50 x 50 cm (middle to ventral part of hippocampus) through to covering a complete environment (168 x 168 cm) (Jung, Wiener, & McNaughton, 1994). Albeit continuous firing fields are not necessarily homogeneously spreading the entire space considered. Hollup, Molden, Donnett, Moser and Moser (2001) found higher place cell density and higher overlap at the hidden platform in a water maze. Hence, multiple place cells in concert seem to be able to capture and represent the significance of an area larger than every single cell alone.

This finding relates nicely to a more perceptual and psychological understanding of a place as a rather discrete entity. Mallot and Lancier (2018), for example, express their notion of a place as follows: “spatial memory seem[s] to be organized in terms of discrete places, which we remember, assign functions to, or communicate about. These places are not geometrical points, but have a spatial extent within which we can perform small movements without always thinking of the newly acquired position as a separate place” (p. 291). Many of the following concepts fit into this description and define places to be formed based on an areas perceptual distinctiveness and/or psychological value.

In order to make a robot navigate through space Hübner and Mallot (2007) differentiate catchment areas and confusion areas in the robots surrounding. When a goal is visible and identifiable from a distance and enables beacon-based homing towards it then one is located in its catchment area. The confusion area is reached close to the goal when the memorized visual cues correspond to the current visual input (i.e., matching snapshots). Then one has reached the goal. The size of this area is depending on the distinctiveness of cues available in memory and at the goal. Thus, a place is an anchor point defined by perceptually salient aspects together with its extended neighborhood—a concept adaptable to human navigation as well. Correspondingly, theories have been presented that are mainly based on encountered views (e.g., Schölkopf & Mallot, 1995). In a view graph, nodes represent views and edges indicate a movement and/or head rotation to encounter the next view, thus reflecting the immediate temporal sequence during exploration of the space. These views can then additionally be connected to a place node.

Poucet (1993) distinguished local charts from places. He proposes “A local chart may contain any number of place representations, but, by definition, all of these place representations must share a number of common stimulus elements provided by either the proximal or distal environment” (p.170). By doing so he takes into account the issue of scale and the hierarchical character of the attempt to define and describe a place based on visual anchor points (i.e., landmarks). In an open field containing multiple local landmarks that are surrounded by distant landmarks such as mountains, each landmark, and its immediate neighborhood can be considered a place, but also the entire field that is defined by its surround.

According to Meilinger (2008), vista spaces serve as local memory units. Vista spaces have been defined for example by Montello (1993) to describe spaces that are visually apprehensible from a single vantage point. However, as we navigate this notion of vista space becomes a continuum. So, what differentiates the moment a navigator is standing in the middle of a corridor from the moment he is standing at the connection between two corridors? Space syntax, a set of architectural analysis techniques (e.g., Hillier, 1996), allows analyzing environmental spaces in terms of possible view axis from any point within the environment. A street or corridor is accompanied with a longer view from one to the other end of the street. Crossroads or turns typically involve the intersection of long view axes (one from each street) rendering these locations plausible transition points between two vista spaces. Additionally, crossroads and turns are typically evoking stronger changes in the percept compared to simply walk straight along a street.

So far, these discretizing notions of places concentrated on visual aspects of a scene. However, there are studies indicating that also the psychological value of a location affects how it is spatially represented. Decision points such as intersections have been found to have a special standing in spatial memory. Landmarks at decision points are faster recognized (e.g., Janzen, 2006; Janzen & Weststeijn, 2007), more often used in route descriptions (e.g., Michon & Denis, 2001; Tversky & Lee, 1999) and elicit higher activity in the parahippocampal gyrus (e.g., Janzen & van Turennout, 2004; Janzen & Weststeijn, 2007) compared to landmarks along a street. Janzen (2006) also found that the route direction effect is primarily found for landmarks at intersections. Due to their high saliency decision point are likely to be represented as place nodes in a graph representation. They are particularly informative as they provide information about connectivity and much more distal cues.

Interestingly, while the vista space approach emphasizes the importance of clearly defined enclosed spaces, such as individual corridors or streets, the above-mentioned findings highlight the importance of decision points, hence, the intersection points between two vista spaces. The high correspondence between the results in all my studies, indeed, indicates that both can serve as molds for local memory units in a graph representation. Both "types" of places, the vista spaces used in study 1, 3 and 4 as well as the intersections in study 2, yielded place-to-place distance effects (study 1, 2, 4) and did not abstract from the directedness of the successive encounter during learning (study 1, 2). Further, in study 2 and 3, the chunks of interrelated trials querying neighboring targets successively both led to the absence of corridor/place distance effects after the first target direction was estimated. Hence, in both studies, subsequent pointings to a neighboring target were probably based on the previous estimate. In sum, independent of the "type" of places used it seems that the same memory structures emerge, and the same processes are at work during recall.

A potential follow-up experiment trying to differentiate continuous and discrete representations of a place and in the case of discrete representations aiming to distinguish between enclosed vista spaces and decision points as potential place candidates could leverage the "place" distance effects found in the presented studies. For example, imagine an environmental space consisting of several interconnected corridors each of which is identified by a distinct color or wallpaper. After learning, participants are teleported to different locations in that space and must point to the middle of predefined corridors various corridors away. By varying participants standpoints within a corridor during testing, it is possible to ascertain when latency increase occurs. If the corridor distance effect is constant irrespective of whether people are standing at the beginning, in the middle or at the end of a corridor, this corridor is likely to have served as an individual memory unit. In contrast, if we find that the locations within the corridor also affects pointing latency this might either speak for a discretization of the environment into multiple subspaces (e.g., turns and middle of corridor) or indeed a continuous mapping of places (see footnote 8). Further variations could involve having people additionally point to the turns connecting two corridors to assess whether these turns serve as memory units as well.

Taken together, the question "What is a place?" cannot be unambiguously answered here. Twilight zones exist which we might be able to enlighten with the help of more sophisticated studies. Nonetheless, these twilight zones might also simply remain due to the high flexibility of humans in selecting places from whatever spatial information are available and of significance to them whilst learning. Importantly, the results of my studies are in accordance with the view that places are represented as distinct memory units, probably in the form of local reference frames, and connected to other places in a network structure that must be accessed successively along the path of connections that are stored.

4 Conclusion

In my thesis, I was able to demonstrate that human survey performance does not rely on an embedded map-like memory structure that specifies metric relations between all visited places in a globally coherent format. Instead, my findings indicate that local places are basic memory units (i.e., nodes) that are forming an interconnected graph structure enriched with local metrics. This graph structure is likely driven by opaque barriers that occlude the view between relevant places, which makes this memory format for environmental spaces fundamentally different from object-to-object relations memorized in a single vista space. Indeed, my results suggest that these place units are composed of locally confined reference frames that are connected to their neighboring place units by directed links specifying the local metrics of how to get from A to B, reflecting the sequential egocentric learning experience. This memory structure determines that survey relations between distant places cannot simply be read-out right away but must be incrementally constructed in a time-consuming process on the fly, thereby following the place-to-place connectivity along the graph. Albeit generally transient, for a short amount of time previous survey estimates can be used to base subsequent estimates on. Importantly, this enriched graph must not be globally embedded in a coherent format. Instead, local place-to-place metrics can be distorted independently of each other, leading to globally impossible representations without inducing conflict. In contrast to learning in a single vista space, we usually do not have consistent access to stable landmarks in environmental space, making it particularly difficult to remain oriented relative to a global main orientation while navigating. However, following a deduction that integrated the findings of all my four studies, enriched graphs might possess a general reference direction, a "mental north" that propagates across multiple local place units in order to coordinate and ease recall thereof. These general reference directions can comprise only a subset of places—thereby allowing to form clusters of vista spaces—or the entire environment that was learned. The use of these general reference directions seems to be flexible and depending on the location of the target.

The findings of my thesis have relevance not only for the ongoing theoretical discussions about the structure of environmental survey knowledge in the spatial cognition community but can also be transferred to a range of more applied problems. The fact that clustering and compartmentalizing of space seem to likewise cluster our memory is affecting us every day in our life. If you need to decide whether you want to tear down the wall between kitchen and living room during renovation to have an open-plan kitchen. If a city needs to decide whether to build a new, view-blocking apartment building on the yet undeveloped area between the city center and outskirts. If you need to describe a route to a city site, wondering whether you should lead the non-local visitor along a more direct path through a cluttered and complex neighborhood or have him make a detour involving longer, straighter streets and fewer turns. Such decisions will alter how you represent your flat, affect how pedestrians plan routes and travel through a city, have an impact on how well the non-local visitor finds his way. Many of these things are known or already implemented for city planning or route descriptions. My thesis shows they are right to do so.

Nowadays we are frequently exposed to maps of our environment. Google maps leads us through every new city or unknown neighborhood. You can easily access maps from any site you want to explore, for example, when visiting a new University campus or when strolling through a museum or exhibition. Larger buildings display emergency exit maps on every floor. The existence of a Euclidean mental map, therefore, has some compelling face validity for us as we are exposed to them ever so often. However, my findings show that there seems to be no coherent map in our head. Despite our elusive imagination that we have a coherent, consistent picture of the world around us in our heads, our memory of it is in fact like an impossible puzzle consisting of imperfect pieces that we need to constantly put together to find our way.

5 Statement of Contribution

This thesis is presented in the form of a semi-cumulative thesis presenting work that is, at the time of submission, either published or in preparation. In the following individual author contributions to the experiments and publications are given for each study.

Meilinger, T., Strickrodt, M., & Bülthoff, H.H. (2016). Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation. *Cognition, 155,* 77-95. doi 10.1016/j.cognition.2016.06.003

- Meilinger, T. and candidate Strickrodt, M. share first authorship
- Meilinger, T.: definition of research problem, study design, programming experiment, data collection, writing and revising paper
- **Candidate Strickrodt, M.:** definition of research problem, data collection, data analysis, writing and revising paper
- Bülthoff, H. H.: revising paper

Meilinger, T., Strickrodt, M., & Bülthoff, H.H. (2018). Spatial Survey Estimation Is Incremental and Relies on Directed Memory Structures. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI. 11th International Conference, Spatial Cognition 2018. Lecture Notes in Artificial Intelligence* (Vol. 11034). Springer. doi 10.1007/978-3-319-96385-3_3

- Meilinger, T.: definition of research problem, study design, programming experiment, data collection, writing and revising paper
- **Candidate Strickrodt, M.:** definition of research problem, writing and revising paper
- Bülthoff, H. H.: revising paper

Strickrodt, M., Meilinger, T., Bülthoff, H.H., & Warren, W.H. (in preparation). Navigators learn a local graph, not a global map, of a complex environment.

- **Candidate Strickrodt, M.:** definition of research problem, study design, programming experiment, data collection, data analysis, writing and revising paper
- Meilinger, T.: definition of research problem, study design, revising paper
- Bülthoff, H. H.: revising paper

- Warren, W.H.: definition of research problem, study design, revising paper

Strickrodt, M., Bülthoff, H.H., & Meilinger, T. (2018). Memory for navigable space is flexible and not restricted to exclusive local or global memory units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi 10.1037/xlm0000624

- **Candidate Strickrodt, M.:** definition of research problem, study design, programming experiment, data collection, data analysis, writing and revising paper
- Bülthoff, H. H.: revising paper
- Meilinger, T.: definition of research problem, study design, revising paper

6 References

Avraamides, M. N., & Kelly, J. W. (2010). Multiple systems of spatial memory: evidence from described scenes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(3), 635–645. http://doi.org/10.1037/a0017040

Beals, R., Krantz, D. H., & Tversky, A. (1968). Foundations of multidimensional scaling. *Psychological Review*, *75*(2), 127–142. http://doi.org/10.1037/h0025470

Bellmund, J. L. S., Deuker, L., Navarro Schröder, T., & Doeller, C. F. (2016). Grid-cell representations in mental simulation. *ELife*, *5*(AUGUST), 1–21. http://doi.org/10.7554/eLife.17089

Brockmole, J. R., & Wang, R. F. (2002). Switching between environmental representations in memory. *Cognition*, *83*(3), 295–316. http://doi.org/10.1016/S0010-0277(02)00012-4

Brockmole, J. R., & Wang, R. F. (2003). Changing perspective within and across environments. *Brain & Development*, *25*(4), 291–293. http://doi.org/10.1016/S0

Burroughs, W. J., & Sadalla, E. K. (1979). Asymmetries in Distance Cognition. *Geographical Analysis*, *11*(4), 414–421. http://doi.org/10.1111/j.1538-4632.1979.tb00709.x

Bush, D., Barry, C., Manson, D., Correspondence, N. B., & Burgess, N. (2015). Using Grid Cells for Navigation Viewpoint Using Grid Cells for Navigation. *Neuron*, *87*, 507–520. http://doi.org/10.1016/j.neuron.2015.07.006

Byrne, P., Becker, S., & Burgess, N. (2007a). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375. http://doi.org/10.1037/0033-295X.114.2.340

Byrne, P., Becker, S., & Burgess, N. (2007b). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375. http://doi.org/10.1037/0033-295X.114.2.340.Remembering

Byrne, R. W. (1979). Memory for Urban Geography. *Quarterly Journal of Experimental Psychology*, *31*(1), 147–154. http://doi.org/10.1080/14640747908400714

Caduff, D., & Timpf, S. (2008). On the assessment of landmark salience for human navigation. *Cognitive Processing*, *9*(4), 249–267. http://doi.org/10.1007/s10339-007-0199-2

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. http://doi.org/10.1016/j.cub.2015.02.037

Chadwick, M. J., Jolly, A. E. J., Amos, D. P., Hassabis, D., & Spiers, H. J. (2015). A Goal Direction Signal in the Human Entorhinal/Subicular Region. *Current Biology*, *25*(1), 87–92. http://doi.org/10.1016/j.cub.2014.11.001

Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review*, *19*(1), 1–23. http://doi.org/10.3758/s13423-011-0182-x

Chrastil, E. R., & Warren, W. H. (2014). From Cognitive Maps to Cognitive Graphs. *PLoS ONE*, *9*(11), e112544. http://doi.org/10.1371/journal.pone.0112544

Cohen, R., Baldwin, L. M., & Sherman, R. C. (1978). Cognitive Maps of a Naturalistic Setting. *Child Development*, *49*(4), 1216. http://doi.org/10.2307/1128763

Derdikman, D., & Moser, E. I. (2011). A Manifold of Spatial Maps in the Brain. In *Space, Time and Number in the Brain* (Vol. 1, pp. 41–57). Elsevier. http://doi.org/10.1016/B978-0-12-385948-8.00004-9

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), 1325–1332. http://doi.org/10.1038/nn.2396

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*. http://doi.org/10.1038/nature08704

Downs, R. M., & Stea, D. (1973). *Image and Environment: Cognitive Mapping and Spatial Behavior*. Chigaco: Aldine Press.

Dragoi, G., & Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, *469*(7330), 397–401. http://doi.org/10.1038/nature09633

Easton, R. D., & Sholl, M. J. (1995). Object-Array Structure, Frames of Reference, and Retrieval of Spatial Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 483–500. http://doi.org/10.1037/0278-7393.21.2.483

Ekstrom, a D., Kahana, M. J., Caplan, J. B., Fields, T. a, Isham, E. a, Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*(6954), 184–188. http://doi.org/10.1038/nature01955.1.

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, *20*(11), 1504–1513. http://doi.org/10.1038/nn.4656

Erdem, U. M., & Hasselmo, M. (2012). A goal-directed spatial navigation model using forward trajectory planning based on grid cells. *European Journal of Neuroscience*, *35*(6), 916–931. http://doi.org/10.1111/j.1460-9568.2012.08015.x

Evans, G. W. (1980). Environmental cognition. *Psychological Bulletin*, *88*(2), 259–287. http://doi.org/10.1037/0033-2909.88.2.259

Farrell, M. J., & Robertson, I. H. (2000). The automatic updating of egocentric spatial relationships and its impairment due to right posterior cortical lesions. *Neuropsychologia*, *38*(5), 585–595. http://doi.org/10.1016/S0028-3932(99)00123-2

Fenton, A. A., Kao, H.-Y., Neymotin, S. A., Olypher, A., Vayntrub, Y., Lytton, W. W., & Ludvig, N. (2008). Unmasking the CA1 Ensemble Place Code by

Exposures to Small and Large Environments: More Place Cells and Multiple, Irregularly Arranged, and Expanded Place Fields in the Larger Space. *Journal of Neuroscience*, *28*(44), 11250–11262. http://doi.org/10.1523/JNEUROSCI.2862-08.2008

Freundschuh, S. M., & Egenhofer, M. J. (1997). Human conceptions of spaces: implications for GIS. *Transactions in GIS*, *2*(4), 361–375. http://doi.org/10.1111/j.1467-9671.1997.tb00063.x

Friedman, A., & Brown, N. R. (2000). Reasoning about geography. *Journal of Experimental Psychology: General*, *129*(2), 193–219. http://doi.org/10.1037/0096-3445.129.2.193

Fujita, N., Klatzky, R. L., Loomis, J. M., & Golledge, R. G. (1993). The Encoding-Error Model of Pathway Completion without Vision. *Geographical Analysis*, *25*(4), 295–314. http://doi.org/10.1111/j.1538-4632.1993.tb00300.x

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*(7132), 190–194. http://doi.org/10.1038/nature05601

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Gallistel, C. R., & Cramer, A. E. (1996). Computations on metric maps in mammals: getting oriented and choosing a multi-destination route. *The Journal of Experimental Biology*, *199*(Pt 1), 211–7. Retrieved from https://pdfs.semanticscholar.org/6d4c/3379ded4a00e836228e26501cac37b143f06.pdf

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, *14*(4–8), 704–715. http://doi.org/10.1080/13506280500193719

Gillner, S., & Mallot, H. A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience*, *10*(4), 445–463. http://doi.org/10.1162/089892998562861

Golledge, R. G., Klatzky, R. L., & Loomis, J. M. (1996). Cognitive Mapping and Wayfinding by Adults Without Vision. In J. Portugali (Ed.), *The Construction of Cognitive Maps* (pp. 215–246). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-0-585-33485-1_10

Golledge, R. G., & Spector, A. N. (1978). Comprehending the Urban Environment: Theory and Practice. *Geographical Analysis*, *10*(4), 403–426. http://doi.org/10.1111/j.1538-4632.1978.tb00667.x

Greenauer, N., & Waller, D. (2010). Micro- and macroreference frames: Specifying the relations between spatial categories in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 938–57. http://doi.org/10.1037/a0019647

Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal Replay Is Not a Simple Function of Experience. *Neuron*, *65*(5), 695–705. http://doi.org/10.1016/j.neuron.2010.01.034

Han, X., & Becker, S. (2014). One spatial map or many? Spatial coding of connected environments. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(2), 511–31. http://doi.org/10.1037/a0035259

Hardiess, G., Gillner, S., & Mallot, H. A. (2008). Head and eye movements and the role of memory limitations in a visual search paradigm. *Journal of Vision*, *8*(1), 7. http://doi.org/10.1167/8.1.7

Herman, J. F., & Siegel, A. W. (1975). The Development of Spatial Representations of Large-Scale Environments. In *Advances in Child Development and Behavior* (Vol. 10, pp. 9–55). http://doi.org/10.1016/S0065-2407(08)60007-5

Hillier, B. (1996). *Space is the machine*. Cambridge, UK: Cambridge University Press. Retrieved from http://discovery.ucl.ac.uk/3881/1/SITM.pdf

Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition*, *13*(3), 208–217. http://doi.org/10.3758/BF03197683

Hirtle, S. C., & Mascolo, M. F. (1986). Effect of semantic clustering on the memory of spatial locations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 182–189. http://doi.org/10.1037/0278-7393.12.2.182

Hodgson, E., & Waller, D. (2007). Capacity Limits of Spatial Updating: Online and Offline Updating. In *Annual meeting of the Psychonomic Society*. Retrieved from http://www.users.miamioh.edu/hodgsoep/Publications/2007/Hodgson&Waller_Psychonomics2007.pdf

Hollup, S. A., Molden, S., Donnett, J. G., Moser, M. B., & Moser, E. I. (2001). Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *The Journal of Neuroscience*, *21*(5), 1635–1644. http://doi.org/21/5/1635 [pii]

Holmes, M. C., & Sholl, M. J. (2005). Allocentric Coding of Object-to-Object Relations in Overlearned and Novel Environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1069–1087. http://doi.org/10.1037/0278-7393.31.5.1069

Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like processing of imagined navigation. *Current Biology*, *26*(6), 842–847. http://doi.org/10.1016/j.cub.2016.01.042

Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., … Spiers, H. J. (2014). The hippocampus and entorhinal cortex encode the path and euclidean distances to goals during navigation. *Current Biology*, *24*(12), 1331–1340. http://doi.org/10.1016/j.cub.2014.05.001

Hübner, W., & Mallot, H. A. (2007). Metric embedding of view-graphs. *Autonomous Robots*, *23*(3), 183–196. http://doi.org/10.1007/s10514-007-9040-0

Hurlebaus, R., Basten, K., Mallot, H. A., & Wiener, J. M. (2008). Route Learning Strategies in a Virtual Cluttered Environment. In C. Freksa, N. S. Newcombe, P. Gärdenfors, & S. Wölfl (Eds.), *Spatial Cognition VI. Learning, Reasoning, and Talking about Space* (pp. 104–120). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-87601-4_10

Huttenlocher, J., Hedges, L. V, & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376. Retrieved from http://psycnet.apa.org/journals/rev/98/3/352/

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://doi.org/10.1016/j.cogpsych.2005.08.003

Israel, I., & Berthoz, A. (1989). Contribution of the otoliths to the calculation of linear displacement. *Journal of Neurophysiology*, *62*(1), 247–263. http://doi.org/10.1152/jn.1989.62.1.247

Jacobs, J., Kahana, M. J., Ekstrom, A. D., Mollison, M. V, & Fried, I. (2010). A sense of direction in human entorhinal cortex. *Proceedings of the National Academy of Sciences*, *107*(14), 6487–6492. http://doi.org/10.1073/pnas.0911213107

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X.-X., … Kahana, M. J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience*, *16*(9), 1188–1190. http://doi.org/10.1038/nn.3466

Janzen, G. (2006). Memory for object location and route direction in virtual large-scale space. *The Quarterly Journal of Experimental Psychology*, *59*(3), 493–508. http://doi.org/10.1080/02724980443000746

Janzen, G., & van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience*, *7*(6), 673–677. http://doi.org/10.1038/nn1257

Janzen, G., & Weststeijn, C. G. (2007). Neural representation of object location and route direction: An event-related fMRI study. *Brain Research*, *1165*(1), 116–125. http://doi.org/10.1016/j.brainres.2007.05.074

Jung, M. W., Wiener, S. I., & McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience*, *14*(12), 7347–7356. http://doi.org/10.1523/jneurosci.4759-06.2007

Kelly, J. W., Avraamides, M. N., & Loomis, J. M. (2007). Sensorimotor alignment effects in the learning environment and in novel environments. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(6), 1092–107. http://doi.org/10.1037/0278-7393.33.6.1092

Kelly, J. W., & McNamara, T. P. (2008). Spatial memories of virtual environments: How egocentric experience, intrinsic structure, and extrinsic structure interact. *Psychonomic Bulletin & Review*, *15*(2), 322–327. http://doi.org/10.3758/PBR.15.2.322

Klatzky, R. L. (1998). Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections (pp. 1–17). http://doi.org/10.1007/3-540-69342-4_1

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial Updating of Self-Position and Orientation During Real, Imagined, and Virtual Locomotion. *Psychological Science*, *9*(4), 293–298. http://doi.org/10.1111/1467-9280.00058

Klatzky, R. L., Loomis, J. M., Golledge, R. G., Cicinelli, J. G., Doherty, S., & Pellegrino, J. W. (1990). Acquisition of Route and Survey Knowledge in the Absence of Vision. *Journal of Motor Behavior*, *22*(1), 19–43. http://doi.org/10.1080/00222895.1990.10735500

Kluss, T., Marsh, W. E., Zetzsche, C., & Schill, K. (2015). Representation of impossible worlds in the cognitive map. *Cognitive Processing*, *16*(S1), 271–276. http://doi.org/10.1007/s10339-015-0705-x

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(1), 47–60. http://doi.org/10.1037/0096-1523.4.1.47

Kosslyn, S. M., Pick, H. L., & Fariello, G. R. (1974). Cognitive maps in children and men. *Child Development*, *45*(3), 707–716. http://doi.org/10.1111/1467-8624.ep12147012

Kunz, L., Schroder, T. N., Lee, H., Montag, C., Lachmann, B., Sariyska, R., … Axmacher, N. (2015). Reduced grid-cell-like representations in adults at genetic risk for Alzheimer's disease. *Science*, *350*(6259), 430–433. http://doi.org/10.1126/science.aac8128

Lappe, M., & Rauschecker, J. P. (1995). An illusory transformation in a model of optic flow processing. *Vision Research*, *35*(11), 1619–1631. http://doi.org/10.1016/0042-6989(94)00220-G

Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. a. (1993). Nonvisual navigation by blind and sighted: Assessment of path integration ability. *Journal of Experimental Psychology: General*, *122*(1), 73–91. http://doi.org/10.1037/0096-3445.122.1.73

Mallot, H. A., & Basten, K. (2009). Embodied spatial cognition: Biological and artificial systems. *Image and Vision Computing*, *27*(11), 1658–1670. http://doi.org/10.1016/j.imavis.2008.09.001

Mallot, H. A., & Lancier, S. (2018). Place recognition from distant landmarks: human performance and maximum likelihood model. *Biological Cybernetics*, *112*, 291–303. http://doi.org/10.1007/s00422-018-0751-4

Marchette, S. A., Ryan, J., & Epstein, R. A. (2017). Schematic representations of local environmental space guide goal-directed navigation. *Cognition*, *158*, 68–80. http://doi.org/10.1016/j.cognition.2016.10.005

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2014). Anchoring the neural compass: Coding of local spatial reference frames in human medial parietal lobe. *Nature Neuroscience*, *17*(11), 1598–1606. http://doi.org/10.1038/nn.3834

Martin, J. F., & Thomson, J. A. (1998). Automatic Spatial Updating during Locomotion without Vision. *The Quarterly Journal of Experimental Psychology Section A*, *51*(3), 637–654. http://doi.org/10.1080/713755776

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology*, *18*(1), 87–121. http://doi.org/10.1016/0010-0285(86)90016-2

McNamara, T. P. (2003). How Are the Locations of Objects in the Environment Represented in Memory? In *Spatial Cognition III* (pp. 174–191). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/3-540-45004-1_11

McNamara, T. P., & Diwadkar, V. A. (1997). Symmetry and asymmetry of human spatial memory. *Cognitive Psychology*, *34*(2), 160–190. http://doi.org/10.1006/cogp.1997.0669

McNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 211–227. http://doi.org/10.1037/0278-7393.15.2.211

McNamara, T. P., Sluzenski, J., & Rump, B. (2008). Human Spatial Memory and Navigation. In H. L. Roediger III (Ed.), *Learning and Memory: A Comprehensive Reference* (Vol. 2, pp. 157–178). Oxford: Elsevier. http://doi.org/10.1016/B978-012370509-9.00176-5

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the “cognitive map.” *Nature Reviews Neuroscience*, *7*(8), 663–678. http://doi.org/10.1038/nrn1932

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* (Vol. 5248 LNAI, pp. 344–360). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-87601-4_25

Meilinger, T., Berthoz, A., & Wiener, J. M. (2011). The integration of spatial information across different viewpoints. *Memory & Cognition*, *39*(6), 1042–1054. http://doi.org/10.3758/s13421-011-0088-x

Meilinger, T., & Bülthoff, H. H. (2013). Verbal shadowing and visual interference in spatial memory. *PloS One*, *8*(9), e74177. http://doi.org/10.1371/journal.pone.0074177

Meilinger, T., Frankenstein, J., & Bülthoff, H. H. (2013). Learning to navigate: Experience versus maps. *Cognition*, *129*(1), 24–30. http://doi.org/10.1016/j.cognition.2013.05.013

Meilinger, T., Frankenstein, J., Watanabe, K., Bülthoff, H. H., & Hölscher, C. (2015). Reference frames in learning from maps and navigation. *Psychological Research*, *79*(6), 1000–1008. http://doi.org/10.1007/s00426-014-0629-6

Meilinger, T., Henson, A., Rebane, J., Bülthoff, H. H., & Mallot, H. A. (2018). Humans Construct Survey Estimates on the Fly from a Compartmentalised Representation of the Navigated Environment. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI. 11th International Conference, Spatial Cognition 2018. Lecture Notes in Artificial Intelligence*. Springer.

Meilinger, T., Riecke, B. E., & Bülthoff, H. H. (2013). Local and Global Reference Frames for Environmental Spaces. *Quarterly Journal of*

Experimental Psychology, *67*(3), 542–569. http://doi.org/10.1080/17470218.2013.821145

Meilinger, T., Schulte-Pelkum, J., Frankenstein, J., Berger, D. R., & Bülthoff, H. H. (2015). Global Landmarks Do Not Necessarily Improve Spatial Performance in Addition to Bodily Self-Movement Cues when Learning a Large-Scale Virtual Environment. In M. Imura, P. Figueroa, & B. J. Mohler (Eds.), *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments* (pp. 25–28). Kyoto: Eurographics Association. http://doi.org/10.2312/egve.20151306

Meilinger, T., & Vosgerau, G. (2010). Putting Egocentric and Allocentric into Perspective. In *Lecture Notes in Computer Science* (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6222 LNAI, pp. 207–221). http://doi.org/10.1007/978-3-642-14749-4_19

Michon, P.-E., & Denis, M. (2001). When and Why Are Visual Landmarks Used in Giving Directions? In D. R. Montello (Ed.), *patial Information Theory. COSIT 2001. Lecture Notes in Computer Science, vol 2205* (pp. 292–305). Springer Berlin Heidelberg. http://doi.org/10.1007/3-540-45424-1_20

Mittelstaedt, H. (2000). Triple-loop model of path control by head direction and place cells. *Biological Cybernetics*, *83*(3), 261–270. http://doi.org/10.1007/s004220000169

Mittelstaedt, M.-L., & Mittelstaedt, H. (2001). Idiothetic navigation in humans: estimation of path length. *Experimental Brain Research*, *139*(3), 318–332. http://doi.org/10.1007/s002210100735

Moar, I., & Bower, G. H. (1983). Inconsistency in spatial knowledge. *Memory & Cognition*, *11*(2), 107–113. http://doi.org/10.3758/BF03213464

Moar, I., & Carleton, L. R. (1982). Memory for routes. *The Quarterly Journal of Experimental Psychology Section A*, *34*(3), 381–394. http://doi.org/10.1080/14640748208400850

Montello, D. R. (1993). Scale and multiple psychologies of space. In *Spatial Information Theory* (pp. 312–321). http://doi.org/10.1007/3-540-57207-4_21

Montello, D. R. (1998). A new framework for understanding thr acquisition of spatial knowledge in large-scale environments. In J. M. Egenhofer & G. R. Golledge (Eds.), *Spatial and Temporal Reasoning in Geographic Information Systems* (pp. 143–154). New York: Oxford University Press. http://doi.org/10.1088/1748-6041/6/2/025001

Moser, E. I., Kropff, E., & Moser, M.-B. (2008). Place Cells, Grid Cells, and the Brain's Spatial Representation System. *Annual Review of Neuroscience*, *31*(1), 69–89. http://doi.org/10.1146/annurev.neuro.31.061307.090723

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 162–170. http://doi.org/10.1037//0278-7393.28.1.162

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(1), 142–157. http://doi.org/10.1037/0278-7393.30.1.142

Mou, W., McNamara, T. P., & Zhang, L. (2013). Global frames of reference organize configural knowledge of paths. *Cognition*, *129*(1), 180–193. http://doi.org/10.1016/j.cognition.2013.06.015

Mou, W., Zhao, M., & McNamara, T. P. (2007). Layout geometry in the selection of intrinsic frames of reference from multiple viewpoints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 145–154. http://doi.org/10.1037/0278-7393.33.1.145

Muryy, A., & Glennerster, A. (2018). Pointing Errors in Non-Metric Virtual Environments. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI. 11th International Conference, Spatial Cognition 2018. Lecture Notes in Artificial Intelligence*. http://doi.org/10.1101/273532

Nadel, L. (2013). Cognitive maps. In D. Waller & L. Nadel (Eds.), *Handbook of spatial cognition.* (pp. 155–171). Washington: American Psychological Association. http://doi.org/10.1037/13936-009

Newcombe, N., Huttenlocher, J., Sandberg, E., Lie, E., & Johnson, S. (1999). What do misestimations and asymmetries in spatial judgement indicate about spatial representation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 986–996. http://doi.org/10.1037/0278-7393.25.4.986

Newcombe, N., & Liben, L. S. (1982). Barrier effects in the cognitive maps of children and adults. *Journal of Experimental Child Psychology*, *34*(1), 46–58. http://doi.org/10.1016/0022-0965(82)90030-3

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–175. http://doi.org/10.1016/0006-8993(71)90358-1

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.

O'Malley, M., Bülthoff, H. H., & Meilinger, T. (2014). How to find a shortcut within a city? Mental walk vs. mental model. *56th Conference of Experimental Psychologists* (*TeaP 2014).* Retrieved from http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:2582155

Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D., & Spiers, H. J. (2015). Hippocampal place cells construct reward related sequences through unexplored space. *ELife*, *4*(JUNE2015), 1–17. http://doi.org/10.7554/eLife.06063

Ólafsdóttir, H. F., Carpenter, F., & Barry, C. (2016). Coordinated grid and place cell replay during rest. *Nature Neuroscience*, *19*(6), 792–794. http://doi.org/10.1038/nn.4291

Pantelides, S. N., Kelly, J. W., & Avraamides, M. N. (2016). Integration of spatial information across vision and language. *Journal of Cognitive Psychology*, *28*(2), 171–185. http://doi.org/10.1080/20445911.2015.1102144

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books.

Poucet, B. (1993). Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. *Psychological Review*, *100*(2), 163–182. http://doi.org/10.1037/0033-295X.100.2.163

Presson, C. C., & Montello, D. R. (1988). Points of reference in spatial cognition Stalking the elusive landmark. *British Journal of Developmental Psychology*, *6*(1960), 378–381. http://doi.org/10.1111/j.2044-835X.1988.tb01113.x

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology. General*, *134*(1), 73–92. http://doi.org/10.1037/0096-3445.134.1.73

Reitman, J., & Rueter, H. H. (1980). Organization Revealed by Recall Orders Confirmed by Pauses. *Cognitive Psychology*, *581*, 554–581. Retrieved from https://ac.els-cdn.com/0010028580900201/1-s2.0-0010028580900201-main.pdf?_tid=fb081ae5-acc4-4ec7-96b7-facbcb71dc13&acdnat=1533113033_178b35ba14618d352ee5c36d58cd8fd1

Richard, L., & Waller, D. (2013). Toward a definition of intrinsic axes: The effect of orthogonality and symmetry on the preferred direction of spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1914–1929. http://doi.org/10.1037/a0032995

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition*, *27*(4), 741–750. http://doi.org/10.3758/BF03211566

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1157–1165. http://doi.org/10.1037/0278-7393.15.6.1157

Rieser, J. J., Pick, H. L., Ashmead, D. H., & Garing, A. E. (1995). Calibration of Human Locomotion and Models of Perceptual-Motor Organization. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 480–497. http://doi.org/10.1037/0096-1523.21.3.480

Rossano, M. J., West, S. O., Robertson, T. J., Wayne, M. C., & Chase, R. B. (1999). The acquisition of route and survey kowledge from computer models. *Journal of Environmental Psychology*, *19*(2), 101–115. http://doi.org/10.1006/jevp.1998.0115

Ruddle, R. A., Howes, A., Payne, S. J., & Jones, D. M. (2000). Effects of hyperlinks on navigation in virtual environments. *International Journal of Human Computer Studies*, *53*(4), 551–581. http://doi.org/10.1006/ijhc.2000.0402

Ruddle, R. A., & Lessels, S. (2009). The benefits of using a walking interface to navigate virtual environments. *ACM Transactions on Computer-Human Interaction*, *16*(1), 1–18. http://doi.org/10.1145/1502800.1502805

Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Humam Learning and Memory*, *6*(5), 516–528.

Sanders, H., Rennó-Costa, C., Idiart, M., & Lisman, J. (2015). Grid Cells and Place Cells: An Integrated View of their Navigational and Memory Function. *Trends in Neurosciences*, *38*(12), 763–775. http://doi.org/10.1016/j.tins.2015.10.004

Saunders, J. A., & Backus, B. T. (2006). Perception of surface slant from oriented textures. *Journal of Vision*, *6*(9), 3. http://doi.org/10.1167/6.9.3

Schick, W., Halfmann, M., & Mallot, H. A. (2015). How to construct a linguistic landmark: language cues in the formation of hierarchical representations of space. *Cognitive Processing*, *16*(1), 383–388. http://doi.org/10.1007/s10339-015-0722-9

Schinazi, V. R., & Epstein, R. A. (2010). Neural correlates of real-world route learning. *NeuroImage*, *53*(2), 725–735. http://doi.org/10.1016/j.neuroimage.2010.06.065

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. http://doi.org/10.1002/hipo.22111

Schölkopf, B., & Mallot, H. A. (1995). View-Based Cognitive Mapping and Path Planning. *Adaptive Behavior*, *3*(3), 311–348. http://doi.org/10.1177/105971239500300303

Schweizer, K., Herrmann, T., Janzen, G., & Katz, S. (1998). The Route Direction Effect and its Constraints. In *Spatial Cognition* (Vol. 15, pp. 19–38). http://doi.org/10.1007/3-540-69342-4_2

Shelton, A. L., & McNamara, T. P. (2001). Systems of Spatial Reference in Human Memory. *Cognitive Psychology*, *43*(4), 274–310. http://doi.org/10.1006/cogp.2001.0758

Shelton, A. L., & McNamara, T. P. (2004). Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 158–170. http://doi.org/10.1037/0278-7393.30.1.158

Sherman, R. C., Croxton, J., & Smith, M. (1979). *Movement and Structure As Determinants of Spatial Representations perhaps axiomatic that knowledge of physical separation in*. *Journal of Nonverbal Behavior* (Vol. 4). Retrieved from https://link.springer.com/content/pdf/10.1007%2FBF00986910.pdf

Sholl, M. J. (2001). The Role of a Self-Reference System in Spatial Navigation. In D. R. Montello (Ed.), *Cosit 2001* (pp. 217–232). Springer Berlin Heidelberg. http://doi.org/10.1007/3-540-45424-1_15

Sholl, M. J., Kenny, R. J., & DellaPorta, K. A. (2006). Allocentric-heading recall and its relation to self-reported sense-of-direction. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *32*(3), 516–533. Retrieved from http://psycnet.apa.org/fulltext/2006-06644-006.html

Sholl, M. J., & Nolin, T. L. (1997). Orientation specificity in represenations of place. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1494–1507. Retrieved from http://psycnet.apa.org/fulltext/1997-38633-013.pdf

Siegel, A. W., & White, S. H. (1975). The Development of Spatial Representations of Large-Scale Environments. *Advances in Child Development and Behavior*, *10*, 9–55. http://doi.org/10.1016/S0065-2407(08)60007-5

Spiers, H. J., Hayman, R. M. a, Jovalekic, A., Marozzi, E., & Jeffery, K. J. (2015). Place Field Repetition and Purely Local Remapping in a Multicompartment Environment. *Cerebral Cortex*, *25*(1), 10–25. http://doi.org/10.1093/cercor/bht198

Steck, S. D., & Mallot, H. A. (2000). The Role of Global and Local Landmarks in Virtual Environment Navigation. *Presence: Teleoperators and Virtual Environments*, *9*(1), 69–83. http://doi.org/10.1162/105474600566628

Stern, E., & Leiser, D. (2010). Levels of Spatial Knowledge and Urban Travel Modeling. *Geographical Analysis*, *20*(2), 140–155. http://doi.org/10.1111/j.1538-4632.1988.tb00172.x

Sternberg, S. (1969). Memory-scanning: mental processes revealed by reaction-time experiments. *American Scientist*, *57*(4), 421–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5360276

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, *10*(4), 422–437. http://doi.org/10.1016/0010-0285(78)90006-3

Street, W. N., & Wang, R. F. (2014). Differentiating spatial memory from spatial transformations. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(2), 602–8. http://doi.org/10.1037/a0035279

Strickrodt, M., O'Malley, M., & Wiener, J. M. (2015). This place looks familiar—How navigators distinguish places with ambiguous landmark objects when learning novel routes. *Frontiers in Psychology*, *6*(1936), 1–12. http://doi.org/10.3389/fpsyg.2015.01936

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). *Head-Direction Cells Recorded from the Postsubiculum in Freely Moving Rats. II. Effects of Environmental Manipulations*. *The Journal of Neuroscience* (Vol. 70). Retrieved from http://www.jneurosci.org/content/jneuro/10/2/436.full.pdf

Taylor, H. A., Naylor, S. J., & Chechile, N. A. (1999). Goal-specific influences on the representation of spatial perspective. *Memory & Cognition*, *27*(2), 309–319. http://doi.org/10.3758/BF03211414

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, *14*(4), 560–589. http://doi.org/10.1016/0010-0285(82)90019-6

Tlauka, M., Carter, P., Mahlberg, T., & Wilson, P. N. (2011). The first-perspective alignment effect: The role of environmental complexity and familiarity with surroundings. *The Quarterly Journal of Experimental*

Psychology, *64*(11), 2236–2250. http://doi.org/10.1080/17470218.2011.586710

Tobler, W. R. (1976). The geometry of mental maps. *Spatial Choice and Spatial Behavior*, (July), 69–81. Retrieved from http://geog.ucsb.edu/~tobler/publications/pdf_docs/cartography/Mental Maps.pdf

Tolman, E. C. (1948). Cognitive maps in rats and men. *The Psychological Review*, *55*(4), 189–208. Retrieved from http://psychclassics.yorku.ca/Tolman/Maps/maps.htm

Tversky, B. (1981). Distortions in Memory for Maps. *COGNITIVE PSYCHOLOGY*, *13*, 407–433. Retrieved from https://www.tc.columbia.edu/faculty/bt2158/faculty-profile/files/1981_Tversky_Distortionsinmemoryformaps.pdf

Tversky, B., & Lee, P. U. (1999). Pictorial and verbal tools for conveying routes. In *Lecture Notes in Computer Science* (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 1661, pp. 51–64). Springer. http://doi.org/10.1007/3-540-48384-5_4

Uttal, D. H., Friedman, A., Hand, L. L., & Warren, C. (2010). Learning fine-grained and category information in navigable real-world space. *Memory & Cognition*, *38*(8), 1026–1040. http://doi.org/10.3758/MC.38.8.1026

Valiquette, C., & McNamara, T. P. (2007). Different mental representations for place recognition and goal localization. *Psychonomic Bulletin & Review*, *14*(4), 676–680. http://doi.org/10.3758/BF03196820

Van Den Berg, A. V., & Brenner, E. (1994). Why two eyes are better than one for judgements of heading. *Nature*, *371*(6499), 700–702. http://doi.org/10.1038/371700a0

Vass, L. K., & Epstein, R. A. (2013). Abstract Representations of Location and Facing Direction in the Human Brain. *Journal of Neuroscience*, *33*(14), 6133–6142. http://doi.org/10.1523/JNEUROSCI.3873-12.2013

Waller, D., & Greenauer, N. (2007). The role of body-based sensory information in the acquisition of enduring spatial representations. *Psychological Research*, *71*, 322–332. http://doi.org/10.1007/s00426-006-0087-x

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(4), 867–882. http://doi.org/10.1037/0278-7393.32.4.867

Wan, X., Wang, R. F., & Crowell, J. a. (2013). Effects of Basic Path Properties on Human Path Integration. *Spatial Cognition & Computation*, *13*(1), 79–101. http://doi.org/10.1080/13875868.2012.678521

Wang, R. F. (2016). Building a cognitive map by assembling multiple path integration systems. *Psychonomic Bulletin & Review*, *23*(3), 692–702. http://doi.org/10.3758/s13423-015-0952-y

Wang, R. F., & Brockmole, J. R. (2003a). Human navigation in nested environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 398–404. http://doi.org/10.1037/0278-7393.29.3.398

Wang, R. F., & Brockmole, J. R. (2003b). Simultaneous spatial updating in nested environments. *Psychonomic Bulletin & Review*, *10*(4), 981–986. http://doi.org/10.3758/BF03196562

Wang, R. F., Crowell, J. A., Simons, D. J., Irwin, D. E., Kramer, A. F., Ambinder, M. S., … Hsieh, B. B. (2006). Spatial updating relies on an egocentric representation of space: Effects of the number of objects. *Psychonomic Bulletin & Review*, *13*(2), 281–286. http://doi.org/10.3758/BF03193844

Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, *77*(3), 215–250. http://doi.org/10.1016/S0010-0277(00)00105-0

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. http://doi.org/10.1016/j.cognition.2017.05.020

Weber, K. D., Fletcher, W. A., Gordon, C. R., Jones, G. M., & Block, E. W. (1998). Motor learning in the "podokinetic" system and its role in spatial orientation during locomotion. *Experimental Brain Research*, *120*(3), 377–385. http://doi.org/10.1007/s002210050411

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. http://doi.org/10.1037/a0035261

Werner, S., & Schmidt, K. (1999). Environmental reference systems for large-scale spaces. *Spatial Cognition and Computation*, *1*, 447–473. http://doi.org/10.1023/A:1010095831166

Wiener, J. M., Kmecova, H., & de Condappa, O. (2012). Route repetition and route retracing: effects of cognitive aging. *Frontiers in Aging Neuroscience*, *4*, 1–7. http://doi.org/10.3389/fnagi.2012.00007

Wiener, J. M., & Mallot, H. A. (2003). "Fine-to-Coarse" Route Planning and Navigation in Regionalized Environments. *Spatial Cognition & Computation*, *3*(4), 331–358. http://doi.org/10.1207/s15427633scc0304_5

Wiener, J. M., & Mallot, H. A. (2006). Path Complexity Does Not Impair Visual Path Integration. *Spatial Cognition & Computation*, *6*(4), 333–346. http://doi.org/10.1207/s15427633scc0604_3

Wiener, J. M., Schnee, A., & Mallot, H. A. (2004). Use and interaction of navigation strategies in regionalized environments. *Journal of Environmental Psychology*, *24*, 475–493. http://doi.org/10.1016/j.jenvp.2004.09.006

Wilson, P. N., Wilson, D. a, Griffiths, L., & Fox, S. (2007). First-perspective spatial alignment effects from real-world exploration. *Memory & Cognition*, *35*(6), 1432–44. http://doi.org/10.3758/BF03193613

Winter, S. S., & Taube, J. S. (2014). Head Direction Cells: From Generation to Integration. In *Space,Time and Memory in the Hippocampal Formation* (pp. 83–106). Vienna: Springer Vienna. http://doi.org/10.1007/978-3-7091-1292-2_4

Wolbers, T., & Wiener, J. M. (2014). Challenges for identifying the neural mechanisms that support spatial navigation: the impact of spatial scale. *Frontiers in Human Neuroscience*, *8*, 1–12. http://doi.org/10.3389/fnhum.2014.00571

Yamamoto, N., & Shelton, A. L. (2005). Visual and proprioceptive representations in spatial memory. *Memory & Cognition*, *33*(1), 140–150. http://doi.org/10.3758/BF03195304

Yamamoto, N., & Shelton, A. L. (2007). Path information effects in visual and proprioceptive spatial learning. *Acta Psychologica*, *125*(3), 346–360. http://doi.org/10.1016/j.actpsy.2006.09.001

Yamamoto, N., & Shelton, A. L. (2009). Sequential versus simultaneous viewing of an environment: Effects of focal attention to individual object locations on visual spatial learning. *Visual Cognition*, *17*(4), 457–483. http://doi.org/10.1080/13506280701653644

Zetzsche, C., Wolter, J., Galbraith, C., & Schill, K. (2009). Representation of space: Image-like or sensorimotor? *Spatial Vision*, *22*(5), 409–424. http://doi.org/10.1163/156856809789476074

Zhao, H., & Warren, W. H. (2015). On-line and model-based approaches to the visual control of action. *Vision Research*, *110*(PB), 190–202. http://doi.org/10.1016/j.visres.2014.10.008

Zhao, M., & Warren, W. H. (2015). How You Get There From Here. *Psychological Science*, *26*(6), 915–924. http://doi.org/10.1177/0956797615574952

7 Full studies

This chapter presents the authors' work that has been described and discussed in this thesis. The full published article or the article in preparation and the supplementary material thereof are included.

7.1 Study 1: Vista vs. Environmental Space

Study reference

Title: Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation

Authors: Tobias Meilinger †, Marianne Strickrodt †, Heinrich H. Bülthoff († shared authorship)

Status: Accepted, Published

How to cite: Meilinger, T., Strickrodt, M., & Bülthoff, H.H. (2016). Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation. *Cognition, 155,* 77-95.

Abstract

Two classes of space define our everyday experience within our surrounding environment: vista spaces, such as rooms or streets which can be perceived from one vantage point, and environmental spaces, for example, buildings and towns which are grasped from multiple views acquired during locomotion. However, theories of spatial representations often treat both spaces as equal. The present experiments show that this assumption cannot be upheld. Participants learned exactly the same layout of objects either within a single room or spread across multiple corridors. By utilizing a pointing and a placement task we tested the acquired configurational memory. In Experiment 1 retrieving memory of the object layout acquired in environmental space was affected by the distance of the traveled path and the order in which the objects were learned. In contrast, memory retrieval of objects learned in vista space was not bound to distance and relied on different ordering schemes (e.g., along the layout structure). Furthermore, spatial memory of both spaces differed with respect to the employed reference frame orientation. Environmental space memory was organized along the learning experience rather than layout intrinsic structure. In Experiment 2 participants memorized the object layout presented within the vista space room of Experiment 1 while the learning procedure emulated environmental space learning (movement, successive object

presentation). Neither factor rendered similar results as found in environmental space learning. This shows that memory differences between vista and environmental space originated mainly from the spatial compartmentalization which was unique to environmental space learning. Our results suggest that transferring conclusions from findings obtained in vista space to environmental spaces and vice versa should be made with caution.

Keywords: spatial memory; navigation; spatial scale; reference frame; distance; order; vista space; environmental space

Introduction

The ability to remember the location of non-visible targets is essential for a multitude of everyday life tasks, such as communicating the direction to the train station to a non-local person or pointing to a certain cupboard in the kitchen to guide your cooking mate. In order to solve such problems, target locations have to be represented in memory. People have the ability to remember locations in their immediate visible surrounding, i.e., vista space, such as rooms, corridors or open spaces (Montello, 1993). In vista spaces, properties of the surroundings and configuration of objects in space can be perceived from one vantage point by taking a look around. Yet, people are also capable of combining information from several interconnected vista spaces, i.e., an environmental space, such as in buildings or cities (Montello, 1993). Information, in this case, has to be gathered by traversing through and experiencing multiple spaces. Object-to-object relations have to be established mentally, for example, by integrating them into a single reference frame.

Prior studies have already indicated differences between spatial representations acquired in vista and environmental spaces. Firstly, it was found that borders of visibility often determine mental updating of object locations. Namely, locations beyond the currently visible vista space (e.g., locations on a campus) are less likely to be updated compared to locations within the same vista space (e.g., objects in a room) (Avraamides & Kelly, 2010; Kelly et al., 2007; Wang & Brockmole, 2003a, 2003b). Such results suggest that the self-to-object updating process concentrates more on the immediate environment and less on distant targets exceeding the current vista space. Secondly, locations within one vista space unit seem to have a greater degree of "mental closeness" than locations separated by spatial borders. Despite having the same Euclidean distance, the distances between objects is judged as being shorter within a single unit (e.g., room) compared to across units (e.g., to the next room) (Kosslyn et al., 1974; McNamara, 1986; Newcombe & Liben, 1982). Thirdly, switching between distinct environmental representations is costly, which manifests in increased response times (Brockmole & Wang, 2002, 2003, Wang & Brockmole, 2003a, 2003b). Also, memory of environmental spaces can be comprised of multiple, local reference frames, one for each single vista unit of the environmental space (e.g., for each travelled passage of a route) (Meilinger, Riecke, et al., 2013; Werner & Schmidt, 1999). In general, these results suggest that entering a new vista space by passing a visual border strongly affects how we represent the space and that an environmental space is potentially represented segmentally, comprising multiple vista space units.

Importantly, most of these experiments did not control for the amount of information that is needed to be processed within a vista or an environmental space. The number of objects that had to be taken into account and the area that needed to be covered mentally was always larger for the environmental space compared to the vista space, for example when retrieving memory of object location within and beyond the current test room, thus increasing memory load for the environmental space compared to the vista space (e.g., Brockmole & Wang, 2002, 2003; Kosslyn et al., 1974; McNamara, 1986; Newcombe & Liben, 1982). Therefore, effects might at least partially be explained by these differences. In order to match information quantity, we examined participants' configurational memory after learning *exactly the same* object layout (keeping distances and angles constant) either within a vista space or in an environmental space.[10] In the following we will derive three hypotheses about potential differences that may arise in the spatial representations of the layout. In a second step, we will examine how distinct learning characteristics within vista and environmental spaces may underlie these differences.

Order effects

Learning an environmental space is inevitably temporal. One needs to pass through a discrete vista unit to perceive the next one. Thus, objects are encountered successively in a specific order. Several studies have examined the effects of order during spatial tasks. Results by Strickrodt, O'Malley and Wiener (2015) suggest that when learning a route, people memorize the sequence of encountered landmarks along the way in combination with the corresponding turning direction. Landmark and turning information of the preceding intersection were used to infer the correct direction of turn at the fol-

[10] Studies utilizing vista space learning usually test what they call object-to-object relations (Avraamides & Kelly, 2010; Mou & McNamara, 2002; Yamamoto & Shelton, 2009). Studies exploring navigation and wayfinding in environmental space typically examine object-to-object relations as well, but subsume it under the term survey knowledge (i.e., knowing where a target is located in terms of direction and distance without necessary knowing a route leading there; e.g., Siegel & White, 1975).

lowing decision point. Object order is also used to identify overall route direction, i.e., forward direction or return path (Wiener et al., 2012). How engrained object order is in spatial memory was demonstrated in a priming experiment by Janzen (2006). After learning a route in a large-scale virtual environment containing a range of landmarks, subsequent recognition was faster when participants were primed with a former predecessor landmark, compared to a former successor landmark (see also Schweizer et al., 1998). These results are in line with the assumption that the representation of a route is highly integrated, following a stimulus-response-stimulus pattern that allows memorizing route landmarks as a sequence (e.g., O'Keefe & Nadel, 1978).

These studies all target characteristics of the acquired route knowledge. Interestingly, in addition to the above-mentioned results, route direction was also shown to influence performance in tasks designed to address configurational memory (survey tasks), even though, typically configurational knowledge is thought to be uncoupled from the order of learning. In a study by Moar and Carleton (1982), participants were more accurate in directional and distance judgements to targets along a route when probed in the direction they had previously learned the route than in the opposite direction. For example, performance was better while standing at the location of the first object along the route and pointing to the third object encountered during learning than pointing from the third object to the first object. These results suggest that route direction is preserved within configurational memory and used not only for route tasks, but also for survey tasks. This result only represents an indirect examination of whether object order is incorporated in participants' configurational knowledge when learning takes place in an environmental space. In the current study, however, we aimed for a direct measure by letting participants perform a configurational placement task, where the layout of environmental objects had to be reproduced from memory. We predict that, when learned in environmental space, the reconstruction of objects follows the order in which they were first encountered. This order should be easiest to retrieve and, as a result, most preferred. In contrast, presentation of an object layout in a vista space does not impose a predetermined learning order. All objects are visible at once. Access of configurational memory could be flexible following random order. Alternatively, scanning patterns during learning might influence retrieval. These scanning paths might be random as well, thus, being unique for every participant. There is also evidence for systematic scanning paths of grid layouts along horizontal paths (Gilchrist & Harvey, 2006; Hardiess, Gillner, & Mallot, 2008). In sum, whereas environmental space learning should predetermine one specific order, the order of retrieving configurational memory from vista space should be much more varied.

Distance effects

Following the abovementioned results (Avraamides & Kelly, 2010; Brockmole & Wang, 2002, 2003; Kelly et al., 2007; Kosslyn et al., 1974; McNamara, 1986; Meilinger, Riecke, et al., 2013; Newcombe & Liben, 1982; Wang & Brockmole, 2003a, 2003b; Werner & Schmidt, 1999), a compartmentalized space might cause the mental representation to be compartmentalized as well. Learning an environmental space is highly restricted compared to vista space learning. Vision of the entire space is obstructed, the order connected vista spaces are successively entered is predefined as well as the walking distance between locations along the route. We assume that retrieving spatial information will depend on this predefined structure of space.

There is evidence suggesting that distance information from the learning experience might still be preserved within configurational memory. Thorndyke and Hayes-Roth (1982) reported an increased error in directional and distance judgements dependent on the number of corridors between the participant's current position and target location. One possible explanation for this increasing error with distance could indeed be that during task execution (retrieval process), memory of the environmental space is retrieved successively, along the route from which the environment was experienced from. This might be realized, for example, by mentally walking down the memorized route starting from the current location and approaching the target (P. Byrne, Becker, & Burgess, 2007a; Sanders et al., 2015) or by constructing a mental model of the non-visible parts of the environment corridor-by-corridor from one's current location (Meilinger, 2008). Both theories predict an increase in computational effort for larger distances (route distance or amount of corridors) between current and target location, since spatial information must be activated successively following the encoding procedure. However, providing evidence for the increase of pointing error with travelled distance and not for an increase in pointing latency (Thorndyke & Hayes-Roth, 1982) legitimates an alternative explanation: the accumulation of error during encoding (compare to path integration model of Fujita, Klatzky, Loomis, & Golledge, 1993). For example, corridors might be assumed shorter than they actually are, the angle of turn at an intersection might be encoded as a regular 90° turn, whereas in fact being 80°. On average, this error will be larger the more distance travelled.

We assume that pointing accuracy and latency are indicative of distinct processes of spatial learning. Whereas accuracy might be associated with the encoding process, that is, the precision of memory, latency during pointing relates to the process of retrieval, that is, accessing the memory content (see also Pantelides et al., 2016). The assumption that error and latency do reflect distinct aspects of cognition is used in other literature as well (Prinzmetal, McCool, & Park, 2005; Sternberg, 1969). We expect error accumulation during encoding to be independent of the time needed to retrieve the distorted memory. Therefore, even if a complete, integrated representation of the environment was built, where no additional processing is needed regardless of inter-object distance, the representation itself could be distorted, leading to an error increase with distance. By observing latency we investigate the retrieval process, which in turn should be bound to the structure of spatial representation. Studies examining path integration already demonstrated that both, error as well as latency, increase in a multi segment path completion task with increasing overall path length and/or number of legs (Klatzky et al., 1990; Loomis et al., 1993; but see Wan, Wang, & Crowell, 2013; Wiener & Mallot, 2006). In contrast to these studies, retrieving configurational memory of an object layout for executing survey tasks strongly relies on a long-term representation of space. Evidence regarding latency for retrieving survey knowledge is still missing.

In a single vista space, learning is comparably unrestricted. Relations and distances between to-be-learned objects can be directly perceived in a commonly visible reference space. Typically, no walking path or encoding order is prescribed. When examining direction and distance judgement between targets learned as spread across a fully visible space, McNamara (1986) demonstrated that the accuracy of judgements was sensitive to the Euclidean distance (i.e., the straight-line distance) between two object locations. Whether this is due to an unprecise layout memory or the retrieval process again can only be assessed when analysing latency. Indeed, Kosslyn, Ball and Reiser (1978) found that when learning the positions of landmarks from a map the time to mentally scan from one to another landmark depends on the straight-line distances between them.

In our study, we set out to examine whether and how the structure of environmental and vista space influences the structure of the corresponding representation. We instructed participants to perform a pointing task after learning an object layout either in vista or in environmental space. In case of a segmented, non-integrated representation of the environmental space, retrieval of spatial memory is expected to be successive, following the corridors. This would lead to an increase of pointing latency with increasing corridor distance to the target. Again, in contrast to changes in error size, alteration in latency would explicitly imply the need to adjust processing time in order to solve the task. Access of vista space memory might be affected by the Euclidean distance, facilitating retrieval of objects nearby.

Reference frame orientation

It has been shown that the representation of space is orientation-dependent (for a review see McNamara, 2003). Here, orientation refers to the alignment of the body or visual field with respect to the environment, thus, the perspective onto the environment (independent of the target bearing). After learning the locations of objects within an environment and being subsequently tested for configurational memory, pointing between objects from certain perspectives leads to better performance compared to other perspectives. When learning took place in a vista space often the best pointing performance is shown from the originally learned orientation, i.e., the initial view upon the object layout (often referred to as 0°), compared to novel orientations. Additionally, contra-aligned (180°) and orthogonal orientations (±90°) seem to be retrieved better than oblique orientation (e.g., 45°). Hence, pointing performance usually yields a w-shape, or saw tooth, performance pattern in error and latency along the range of tested body orientations (e.g., Kelly & McNamara, 2008; Meilinger & Bülthoff, 2013; Mou & McNamara, 2002; Shelton & McNamara, 2001). One explanation for this pattern is the encoding of object locations relative to one or two orthogonal reference axes which are retrieved rather effortlessly. Testing from other perspectives requires additional inferential processes (McNamara et al., 2008; Mou et al., 2004; for an alternative explanation of this pattern see Street & Wang, 2014).

The alignment of the spatial reference frame (i.e., orientations on which maximum pointing performance is centred) was found to be influenced by multiple factors. Not only the perspective during encoding (experienced views) is thought to be used, but also environmental geometry (extra-layout cues), such as the shape of the room or the mat on which objects were placed, and the intrinsic configuration of the object layout itself (intra-layout cues) (Mou & McNamara, 2002; Shelton & McNamara, 2001). We assume that for environmental spaces the initial view and the global layout-intrinsic orientation are less determining for setting the reference frame alignment. Deriving the global layout is effortful and cannot be done until the last unit is reached. Instead, each room, corridor and street, constitutes a separate entity, which itself entails discrete intra and extra-layout cues. When walking down, for example, a corridor, the observers' view will naturally become aligned with the geometric axis of the corridor. Such a viewer-space-alignment experienced later during learning was found to be more important for determining reference frame orientation than initial views on a room (Kelly & McNamara, 2008; Shelton & McNamara, 2001; Valiquette & McNamara, 2007). A relatively simple environmental space (e.g., few orthogonally interlinked corridors) may still be represented along a single main axis that spans the entire environmental space (axis presumably aligned with the first vista space encountered; Meilinger, Frankenstein, Watanabe, Bülthoff, & Hölscher, 2015; Tlauka, Carter, Mahlberg, & Wilson, 2011; Wilson, Wilson, Griffiths, & Fox, 2007). However, sufficiently complex environmental spaces seem to be represented within multiple local reference frames, with each local corridor or street occupying a distinct reference frame aligned with the respective corridor (Meilinger, Riecke, et al., 2013; Werner & Schmidt, 1999).

In the current study we kept both the initial view within the environments and the orientation of the global object layout constant while setting the geometric axes of both learning spaces in conflict; this contrasts the reference frame alignment of vista and environmental space learning. We predicted that the reference frame in both environments should be aligned with the visible context. That is the room orientation in vista space and the corridor orientation in environmental space.

The current study is concerned with the acquisition of object-to-object relations under different learning conditions. Experiment 1 examined whether the acquired memory is different depending on whether exactly the same object layout is either learned in vista or in environmental space. We predicted that knowledge acquired from environmental space preserves features of the spatio-temporal learning process, resulting in higher latency when pointing to targets with increasing corridor distance, and recall in the order objects were encountered in. We expected recall latency and order of vista space memory to be influenced by the layout structure instead. Furthermore, we predicted reference frame orientation in environmental space to be aligned with the visible context of the corridor and the learning experience rather than the initial view or the intrinsic layout orientation. In a second step (Experiment 2) we mimicked characteristics of environmental space learning (i.e., movement through space, successive learning experience) within a vista space to isolate the distinguishing factors between spaces.

Experiment 1

Performance in visual pointing and object placement was ascertained after learning an object layout either in an environmental space (ES) or in a vista space (VS).

Method

Participants

26 naïve participants were recruited from a subject database, gave written informed consent and participated in exchange for monetary compensation. Participants were randomly assigned to either of two conditions (ES or VS). Two participants had to be excluded. One participant did not perform significantly better than chance level (90°) in the pointing task. The other participant had to be excluded due to a lack of comprehension of task instructions. The remaining sample of 24 participants (12 for each condition) had a mean age of *M*=26.09 (*SD*=6.94, [19;52]) and included twelve females (seven randomly assigned to ES condition, five to VS condition). The experimental procedure was approved by the ethical committee of the University Hospital Tübingen.

Material

We used Virtools® 5.0 (Dassault Systemes) for programming the virtual environment and the experimental procedure. The experiment took place in a 12x12 m tracking hall, enabling free movement in real space while wearing a head mounted display (HMD) visualizing the virtual space. Participants' head coordinates were tracked by 16 high-speed motion capture cameras with 120 Hz (Vicon® MX 13) to render a real-time egocentric view of the virtual environment. We used a NVIDIA Quadro FX 3700M graphics card with 1024 MB RAM and a nVisor SX111 HMD with a field of view of 102° (horizontal) × 64° (vertical), a resolution of 1280 × 1024 pixels for each eye, and 66% overlap. The interpupillary distance was fixed at 6 cm. We adjusted the HMD fit and screen position for each participant. This virtual reality setup provided important depth cues such as stereo vision, texture gradients, and motion parallax and enabled participant to physically walk through a virtual world.

The object layout participants were asked to learn consisted of seven target objects lying on the floor of the virtual environment arranged within an incomplete 3 × 3 grid with bilateral symmetry (Fig. 1). From left to right a teapot, a hammer and a banana were located in the closest row, and the middle row held a horse, a telephone and a tennis racket. A trumpet was located in the center of the furthest row. It is assumed that the linear relation between hammer, hair dryer and trumpet determines the global main axis of the object layout, namely 0° (compare to Greenauer & Waller, 2010; Kelly & McNamara, 2008). Additional objects by or on the walls, such as a vase or a fireplace, served as aids for orientation within the environment and remained visible throughout learning and testing phase, whereas the target objects were absent in the testing phase. Note that both ES and VS contained the same target objects and objects aiding orientation.

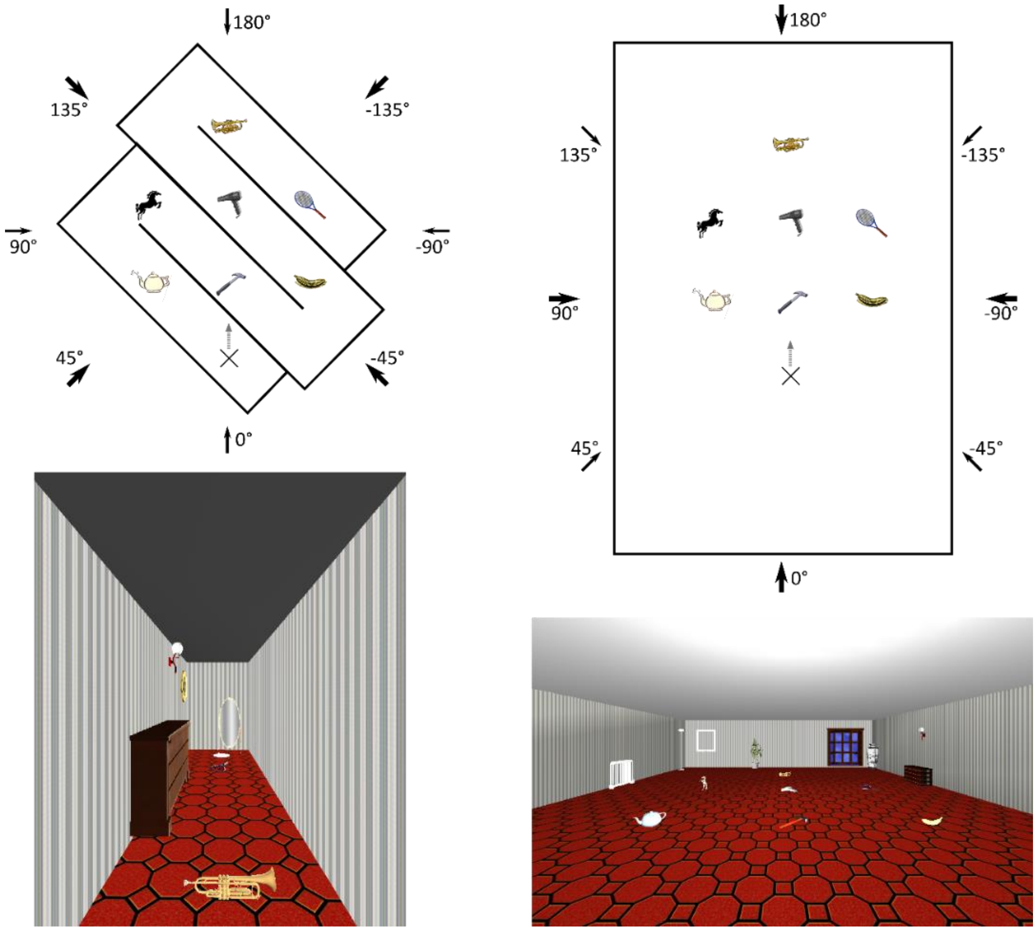


Fig. 1. Left: The layout of ES condition from a birds-eye perspective and participants view from within the environment. Right: Layout and participants view from within the VS condition. Xs indicate the starting position, grey arrows above the Xs the initial view upon the environment. Layout orientation is similar in both environments. Alignment of the visible geometry differs, as indicated by the black, bold arrows.

Both, in ES and VS condition, exactly the same object layout was arranged on the floor, thus, distance and relations between the objects were identical. In the VS condition a rectangular room was presented to the participants from a constant point of view located in front of the object layout (X in Fig. 1, right). Body location within the environment was kept constant during the whole experiment. However, participants were allowed to look around. In the ES condition the environment consisted of walls placed to arrange four parallel, interlinked corridors, offset by 45° to the main axes of object layout and room orientation in the VS condition (Fig. 1, left). To see all objects participants had to walk from the start point X through all corridors sequentially passing each object. Initial view of both environments was set along the main axis of the object layout (grey arrows above the X's in Fig. 1). This view also defined the zero-point of body orientation, as illustrated by the arrow flanked by 0° below the sketches of the environments in Fig. 1, top. Thus, for example, turning 45° to the left in ES in order to be aligned with the corridor would correspond to a body orientation of −45° with respect to the reference orientation of 0°.

Procedure

After participants were familiarized with the equipment the learning phase started. Participants were positioned standing in the corner of the experiment room facing the opposite corner, shortly before being equipped with the HMD. We instructed participants to learn in depth where the virtual objects on the floor were located within the virtual environment and gave no time restrictions or details about later performance tests.

Participants in ES condition moved through the environmental space, following the corridors. They moved twice from their start point, to the end of the last corridor and back to the start to ensure sufficient learning. Since no movement was required in VS condition, participants were instructed to inspect the whole room, including the walls and corners at their back, for later orientation. At no point they were allowed to leave their current position. To ensure correct object identification each object was tagged (in English) by the participant in both conditions. The experimenter corrected misidentifications (i.e., object names not used later in the experiment). After traversing through space twice in ES condition or indicating sufficient learning in VS condition, we removed the target objects and participants proceeded to the learning test. We successively presented blue spheres at former object positions. Participants then had to recollect and name the object located at this position from memory. In ES participants were obliged to walk through the environment again to encounter all blue spheres. The order of testing was the following: Teapot, hammer, hair dryer, racket, trumpet, banana, and horse. This order was neither along rows and columns of the object layout nor along the order of first contact in ES condition. One or more errors resulted in a new learning trial (walking there and back in ES condition and self-paced learning in VS condition). This procedure was repeated until all locations were associated with the correct object. Subsequently, after a short break, the test phase started. The procedure of the test phase was identical for ES and VS condition. Participants stood in front of a table, mounted by a joystick and, first, had to conduct a visual pointing task within their previously learned environment (in the absence of the objects) and, subsequently, perform an object placement task.

In the visual pointing task all target objects were removed from the environment. Participants were teleported to a former object location in each trial, being randomly aligned with 1 out of 8 body orientations. Note that the physical orientation of the participant in the real tracking hall did not change (aligned with the table mounted with the joystick). Rather, for each trial the virtual reality was adjusted in position and orientation to render the desired trial characteristics. The possible orientations participants bodies were then aligned with in the virtual reality are illustrated by the arrows encircling the outline of the environments in Fig. 1, top. They are spaced around a full circle in steps of 45°. The current location (e.g., "You are at the hair dryer") and the pointing target (e.g., "Point to the banana") was indicated during each trial on the HMD screen. This example illustrated in Fig. 2, top, emulated a body orientation of −45° with respect to the reference orientation of 0°. Participants had to identify their orientation based on the visual input from looking around. They were not allowed to walk through the environment during the test phase.

Each of the eight body orientations was tested nine times from different object locations resulting in 72 pointing trials. For 8 of the 9 pointing trials of each body orientation the correct bearings of the target (the correct pointing direction) were spaced around a full circle in steps of 45°. In each body orientation participants had to point to the front, right-front, right, right-back, back, etc. To analyze distance effects, four of the remaining eight trials were set up to cover the largest distance in terms of corridors, i.e., pointing from the teapot to the racket/trumpet and vice versa. This led to correct pointing directions of either 18°, 64°, −108° and −63° relative to the current body orientation. For the remaining four trials targets were set to be located in the same corridor (minimum distance in terms of corridors) with the correct pointing direction of 0°, 45°, −45° and −90°. In summary, 20 trials covered the minimum distance in terms of corridors, 29 trials covered a short distance (next corridor), 19 trials covered a medium distance (second next corridor) and 4 trials covered maximum distance. Trials were presented in random order to every participant. Participants executed pointing by moving a joystick handle, enabling pointing measurements across a 360° circle. For example, assuming the target to be located in front of one's current position, the joystick handle had to be pushed straight forward; assuming the target to be located 135° to the left of one's current position, the joystick handle had to be pulled backwards and to the left. The current pointing direction was accepted with a button press.

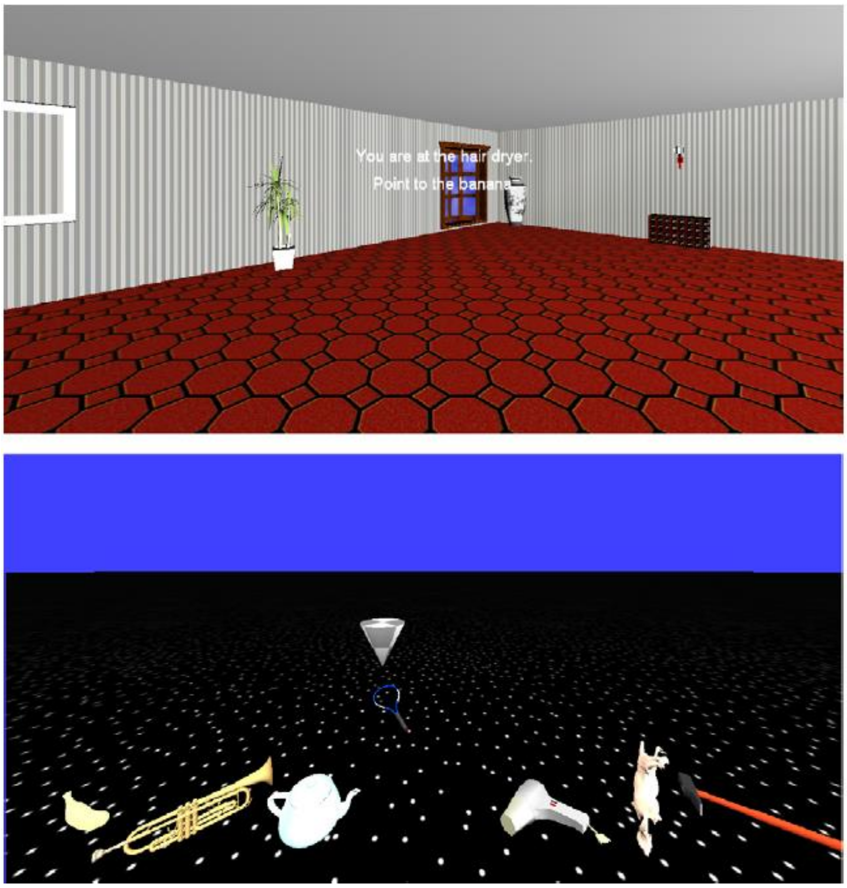


Fig. 2. Top: View upon the environment of one exemplary trial in the pointing task. Bottom: Reconstruction of the layout in the placement task. The currently selected object for rearrangement is marked with an inverted cone. Note that when presenting the environments with the correct field of view no visual distortion was present.

Participants were allowed to examine the environment by looking around. Consequently, during this time head orientation deviated from the body orientation. In order to point to the target, however, participants were instructed to look straight again, thus, realigning with the body orientation in this trial (also realigning with the joystick in front of them). Participants had to align properly in order to have their pointing response recorded and to continue with the next trial. Head orientations deviating more than 10° were not accepted by the program. Latency consisted of the duration between trial onset (appearance within the environment) and button press. Pointing error consisted of the absolute deviation (in °) between pointing direction and correct direction.

For the object placement task, participants were placed in a new virtual surrounding, containing a horizontal plane only (Fig. 2, bottom). Objects were located in a horizontal row in front of the participant, ordered in random sequence determined for each participant. We instructed participants to arrange the objects in the layout they had previously seen in the virtual environment. Objects could be placed in any preferred order and could be (re)arranged until the participant was satisfied. Participants used buttons to switch between objects and the joystick to move the currently selected object. We recorded the order of replacement, i.e., which object was moved away from the start location first, which object was moved second, etc.

Data analysis

From the 24 participants approximately 4% of the pointing performance data was deleted due to deviation of more than two *SD* from a participant's overall mean. All tests conducted were corrected for nonsphericity or inequality of variance when appropriate. As adding participants' gender to the analysis did not change any of the reported effects, we only report the pooled data.

Results

Distance to target

We examined the influence of distance to a target on pointing latency in order to estimate whether this spatial information structured spatial memory. Two potentially meaningful distances between the current position and the target object were analyzed separately. Firstly, the distance in terms of corridors, and secondly, the Euclidean distance. Corridor distance distinguishes trials representing pointing within the same corridor (0), to the next corridor (1), across two (2) or three corridors (3). This classification is based on the ES layout, but was likewise applied to VS trials for control. Euclidean distance represents the straight-line distance between two objects. We differentiate the relative Euclidean distance value of 1, representing the smallest possible distance between two objects (e.g., from teapot to horse) and ascending from there Euclidean distance (e.g., from teapot to hair dryer), 2 (e.g., from teapot to banana) and (e.g., from teapot to trumpet). A relative Euclidean distance value of 1 represents an absolute physically distance of ca. 2.83 m. For analysis it is not meaningful whether absolute or relative values are used. It should be noted here that the two distance dimensions tended to be associated with one another in our study by $r = 0.22$. Also we did not analyze absolute error as this was not the goal of the study. The following analyses were conducted for both distance dimensions separately.

Starting with the dimension of corridor distance and its influence on pointing latency, we conducted an ANOVA with the between-participant factor environment (ES vs. VS) and the within-participant factor corridor distance. A main effect of corridor distance, $F(3, 66) = 5.47$, $p = 0.008$, $\eta_p^2 = 0.20$, and an interaction of corridor distance × environment, $F(3, 66) = 4.58$, $p = 0.017$, $\eta_p^2 = 0.17$, but no main effect of environment, $F < 1$, $p > 0.800$, were found. Thus, the influence of corridor distance on latency differed between environments. To further examine this interaction we regressed pointing latency onto corridor distance separately for each participant. From these regressions b was extracted. This standardized slope describes the linear change of latency with increasing corridor distance for every participant. Fig. 3, top left, depicts the individual and mean slopes for the two environmental conditions. T-tests were used to analyze slopes. As expected, slopes derived from ES did exceed 0, mean $b = 0.22$, $t(11) = 8.15$, $p < 0.001$, $d = 2.35$ (see Fig. 3, top right), indicating an increase in pointing latency the more corridors are residing between current and target location. With each additional corridor pointing took on average 1.02 s ($SD = 0.53$) longer. In VS, which worked as the control condition where no latency increase with ascending corridor distance was expected, the mean slope did not differ from 0, mean $b = 0.02$, $t(11) = 0.63$, $p = 0.542$, $d = 0.18$. Thus, pointing did not take longer the further away targets were located with respect to corridors. Additionally, a comparison of mean slopes between ES and VS revealed that the linear increase of latency across ascending corridor distance was higher for ES compared to VS, $t(22) = -4.28$, $p < 0.001$, $d_s = 1.75$. This pattern was also evident in individual slopes. Whereas in ES data of each single participant rendered a positive slope, in VS this was only the case for 7 out of 12 participants. Results suggest that in ES memory, the spatio-temporal pattern of learning was preserved and this was not due to the structure of the layout itself.

As corridors were not meaningful (since nonexistent) in VS we conducted a control analysis with Euclidean distance. The ANOVA revealed a significant interaction of Euclidean distance × environment, $F(3, 66) = 5.19$, $p = 0.020$, $\eta_p^2 = 0.19$. Even though not significant, Euclidean distance tended to influence pointing latency, $F(3, 66) = 3.13$, $p = 0.073$, $\eta_p^2 = 0.13$. No main effect of environment, $F < 1$, $p > 0.660$ was found. Following this, the effect of Euclidean distance seems to differ between the two environments. Fig. 3, bottom left, depicts the mean and individual slopes derived from the regression of pointing latency onto Euclidean distance. The level of standardized slopes derived from ES did exceed 0, mean $b = 0.08$, $t(11) = 2.57$, $p = 0.026$, $d = 0.74$ (Fig. 3, bottom right), although smaller in size compared to the analysis of corridor distance. In the VS condition, the mean slope did not differ from a 0 slope, mean $b = 0.05$, $t(11) = 1.46$, $p = 0.172$, $d = 0.42$. Thus, only in ES pointing latency increased with increasing Euclidean distance between current and target location. When directly comparing ES and VS slopes did not differ between the environments, $t(22) = -0.75$, $p = 0.470$, $d_s = 0.30$. Looking at individual slopes 8 out of 12 participants (ca. 66%) had a positive slope in the VS condition, 9 out of 12 (75%) in the ES condition. Straight-line distances did not seem to play a prominent role when VS memory was retrieved.

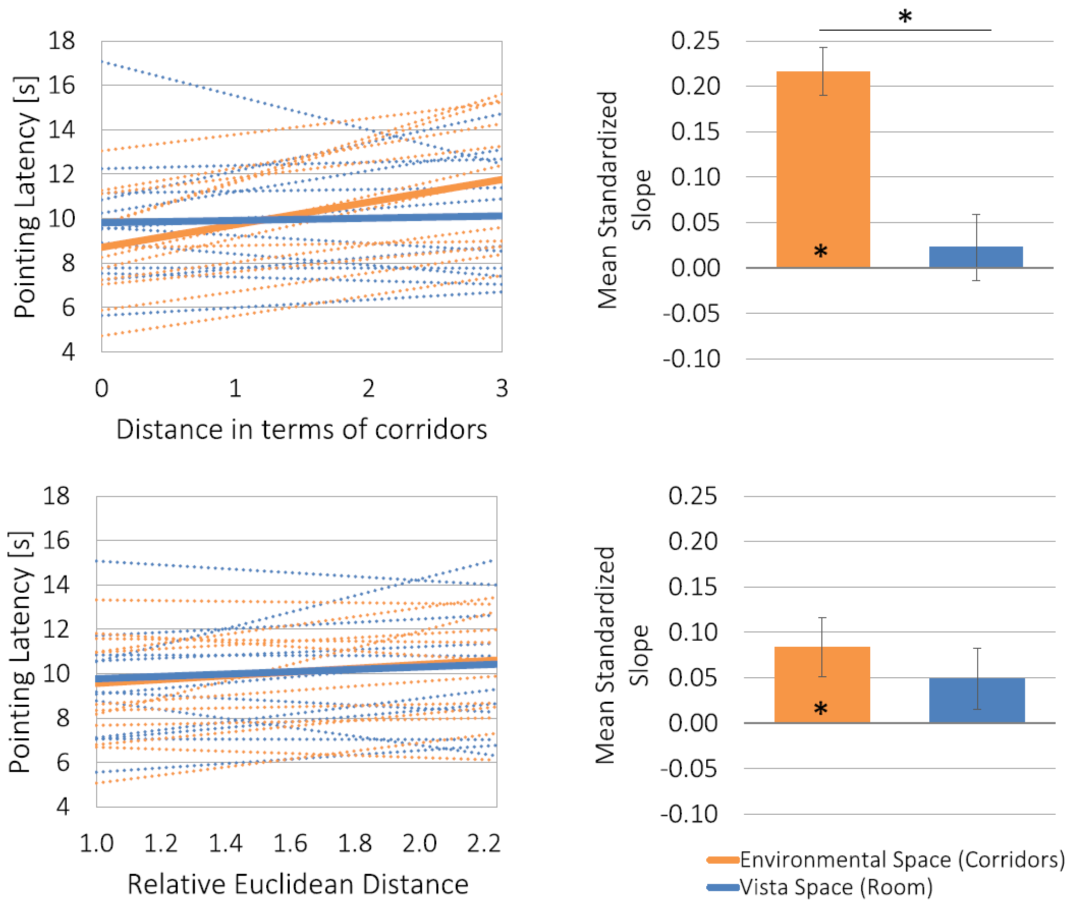


Fig. 3. Top: Linear increase of pointing latency with corridor distance between current position and target. Bottom: Linear effect of Euclidean distance on latency. Left: Individual (dashed line) and averaged (solid line) slope of pointing latency as a function of distance to a target. Right: Mean standardized slopes. Error bars depict SEM. * p<0.05. Asterisks within bars indicate deviation from 0.

Reference frame orientation

To evaluate the reference frame orientation in memory, data derived from the visual pointing task was submitted to a repeated measure ANOVA with the factors environment (ES vs. VS) and body orientation (−135° to 180° in steps of 45°). We expected an interaction. Second, pointing error and latency were fitted to a w-contrast, which describes the primary pattern of performance found in prior research. Using a contrast avoids execution of multiple pairwise comparisons between levels of conditions, but instead makes it possible to describe the fit of a curve to a predefined shape with one single parameter (Greenauer and Waller, 2010, Levin and Neumann, 1999). This predefined shape or pattern has to be specified in advance reflecting your hypothesis. As learning perspective, visible intrinsic object layout, and room orientation in VS condition all entail the same main axis (0° perspective), this value was selected as baseline for the contrast. The utilized w-contrast describes a saw tooth pattern centered on this orientation of 0° (see the legend of Fig. 4, Fig. 5, middle, for illustration). The contrasts weights were defined to be lowest (−1) for the 0° orientation and all orientation which are orthogonal (±90°) or opposed to it (180°), thus, predicting lowest error rate and fastest responses at these orientations. Highest error rate and slowest responses (worst performance) was ascribed to oblique orientations (±45° and ±135°) by setting higher contrasts weights (1). Note that a contrast weight of 0 would predict average performance. To calculate contrast fit to the data, for each participant contrast weights were multiplied with the average performance in the respective perspective and added up (e.g., −1 × average in −180° + 1 × average in −135°, etc.). Contrast fits were inspected using t-tests. We predicted a high, therefore, positive contrast fit for learning an object layout in VS. Here the body orientations leading to best performance should be the initial view upon the environment (0°), and ±90° and 180° deviation from it. For learning in ES setup, however, even though the global layout and the initial view were aligned with 0°, we predicted a reference frame alignment with ±45° and ±135°, according to the visual input when walking through a corridor. A negative w-contrast fit is expected here. A contrast fit of 0 would indicate, that data can't be described by a w-shape.

The underlying assumption of the w-contrast is that the space is represented along two orthogonal axes (four body orientations rendering highest performance, four orientations rendering lowest performance). Alternatively, the space could be represented along a single axis, which can be expressed by an m-contrast centered on either one axis of the corresponding w-contrast. The pattern of an m-contrast implies that pointing performance is best when aligned with this specific axis. Thus, being aligned or directly opposed to one view should then yield best performance whereas decline occurs when deviating from these views, e.g., best performance with 0° and 180° body orientation, worst performance with ±90°. In order to examine which pattern represents our data best, in the last step we tested whether a m-contrast centered on 0° (along long room axis) or a m-contrast centered on ±90° (along short room axis) renders a better data fit than the previously examined w-contrast centered on

0°/±90° for the VS data. Similarly, for ES condition we tested whether the m-contrast centered on 45°/−135° (facing the corridor wall) or the m-contrast centered on −45°/135° (along corridor axis) renders a better data fit than a w-contrast centered on ±45°/±135°. In short, we tested whether the assumption of a single axis fits the data structure better than the assumption of two orthogonal axes. In Fig. 4, Fig. 5, right, w- and m-contrasts of the respective learning condition are depicted. Note that the m-contrast centered on 45°/−135° is the inverse of the m-contrast centered on −45/135°. The same holds for m-contrasts centered on ±90° and 0°/180°. Thus, a positive fit of either corresponds to a negative fit of equal size for the respective other. Regarding the question of best fit, consequently, only positive fits are of interest and were compared via t-tests to the corresponding w-contrast.

Pointing latency

Fig. 4, left, depicts individual (dashed line) and averaged (solid line) pointing latencies of both conditions. The ANOVA yielded a significant main effect of body orientation on pointing latency, $F(7, 154) = 4.50$, $p = 0.002$, $\eta_p^2 = 0.17$, and an interaction of body orientation × environment, $F(7, 154) = 4.93$, $p = 0.001$, $\eta_p^2 = 0.18$, but no main effect of environment, $F < 1$, $p > 0.900$. As predicted, the interaction demonstrates that pointing latency at specific body orientations differed depending on the learning environment. W-contrast fits further identified the nature of these differences (Fig. 4, middle). In line with our hypotheses, w-contrast fit of pointing latency data was higher in VS than in ES, $t(22) = 3.95$, $p = 0.001$, $d_s = 1.61$, suggesting different reference frame orientations. Average values for contrast fit (Fig. 4, middle) were significantly below 0 for ES, $t(11) = -3.48$, $p = 0.005$, $d = 1.00$. This indicates that pointing latency in ES can be well explained by an inverted w-contrast, centered on the oblique ±45° orientations. Even though not significant, a trend was found for the contrast fit for VS to be larger than 0, $t(11) = 1.88$, $p = 0.087$, $d = 0.54$.

Descriptive data of single participants support the pattern of w-contrast fits. In VS 9 out of 12 participants (75%) pointed faster in trials of aligned body orientation (0°, ±90° and 180°) compared to trials of oblique body orientation (±45°, ±135°). In contrast, in ES only 1 out of 12 participants showed this pattern. Thus, the remaining participants (ca. 92%) pointed faster from oblique body orientations.

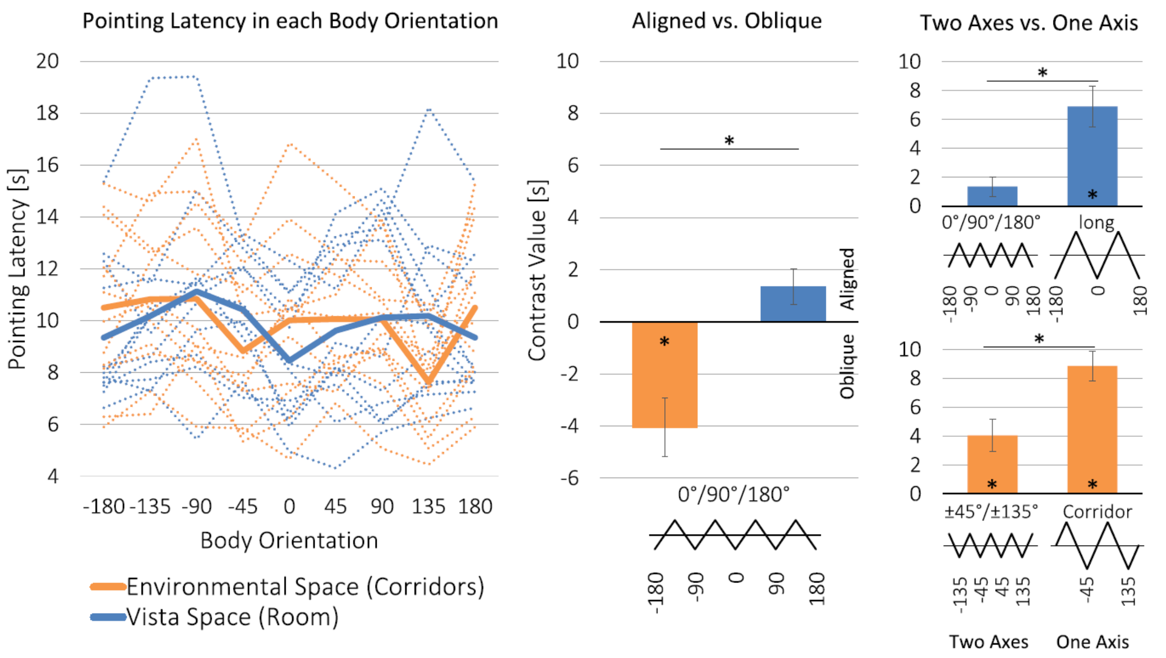


Fig. 4. Left: Individual (dashed line) and averaged (solid line) pointing latency as a function of body orientation. 180° is displayed twice for symmetry. Middle: Values for w-contrast fit centered on 0°/±90°/180° for pointing latency data. Right: Absolute values of contrast fit for w- and m-contrasts, separately for VS (upper) and ES (lower) condition. Pictograms define the used contrasts. Long = along long axis of room. Corridor = along corridor axis.

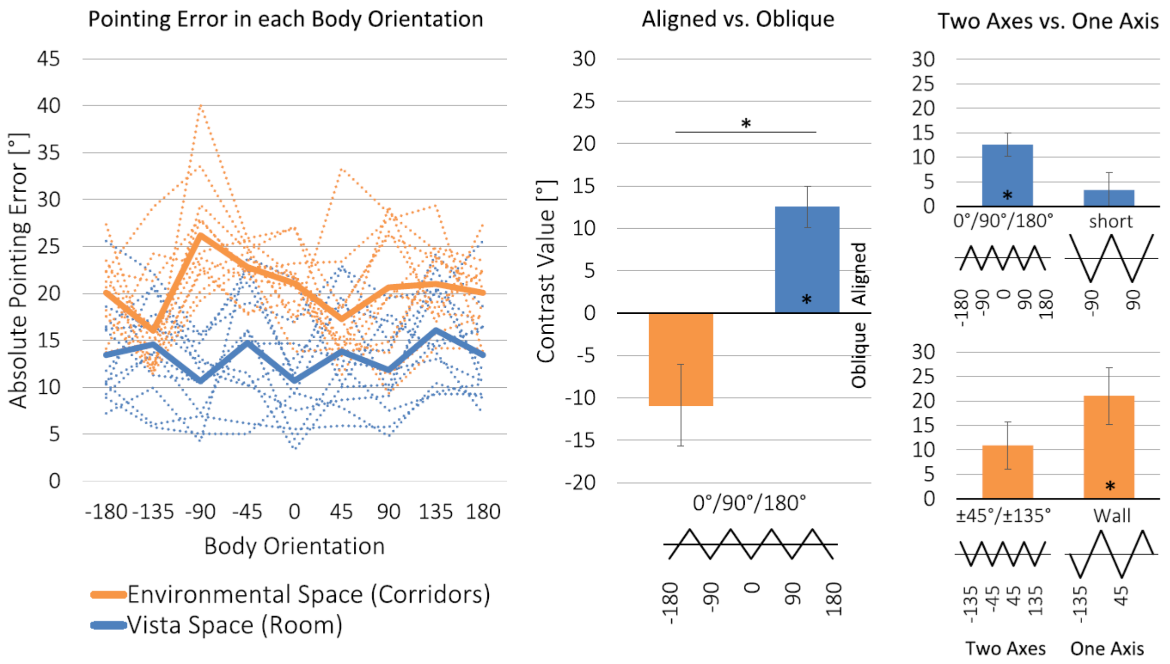


Fig. 5. Left: Individual (dashed line) and averaged (solid line) absolute pointing error. Middle: Mean values for w-contrast fit centered on 0°/±90°/180° for pointing error. Right: Contrast fit for w- and m-contrast, separately for VS (upper) and ES (lower) condition. Pictograms define the center of the contrasts. Short = along short axis of room. Wall = facing corridor wall.

To test whether the assumption of a single axis fits the data structure better than the assumption of two orthogonal axes, data fit to the corresponding m-contrasts was compared to the w-contrast fit for each condition. In Fig. 4, right, w- and m-contrast fits of the respective learning condition are depicted. For VS the positive m-contrast fit centered on 0°/180° (long room axis) significantly exceeded the w-contrast fit centered on 0°/±90°/180°, $t(11) = -3.22$, $p = 0.008$, $d_z = 0.93$. For ES the m-contrast centered on −45/°135° (along corridor axis) produced a positive fit that significantly exceeded the w-contrast fit centered on ±45°/±135°, $t(11) = -4.78$, $p = 0.001$, $d_z = 1.38$. Furthermore, both described m-contrasts exceeded 0 significantly, *t's* > 4.67, *p's* < 0.002. Thus, regarding pointing latency assuming a single reference axis aligned with the longest axis of the visible space fits the data better compared to two orthogonal axes.

Absolute pointing error

Fig. 5, left, depicts individual (dashed line) and averaged (solid line) absolute pointing error of both conditions. The ANOVA revealed a main effect of environment, $F(1, 22) = 27.6$, $p < 0.001$, $\eta_p^2 = 0.556$. Participants pointed more accurately within the VS, indicating an advantage for environmental learning when the object layout was fully visible from one point of view. We also found a main effect of body orientation, $F(7, 154) = 2.75$, $p = 0.010$, $\eta_p^2 = 0.111$, qualified by a significant interaction of body orientation × environment, $F(7, 154) = 6.47$, $p < 0.001$, $\eta_p^2 = 0.23$. Hence, as predicted, error size at specific body orientations differed depending on the learning environment (environmental vs. vista). According to our predictions, w-contrast fit was higher in VS than in ES, $t(22) = 4.14$, $p < 0.001$, $d_s = 1.69$, suggesting differently oriented reference frame orientations (Fig. 5, middle). Moreover, average values for w-contrast fit were above 0 for VS, $t(11) = 4.96$, $p < 0.001$, $d = 1.43$. This indicates that reference frames in VS were oriented along orthogonal directions of room walls and the intrinsic orientation of the object layout (i.e., 0°, ±90° and 180°). In contrast, reference frame orientations in ES tended to be oriented along oblique orientations, $t(11) = -2.15$, $p = 0.055$, $d = 0.62$. Clearly, learning in ES determined a reference frame perspective different from VS learning, even though the object layout was exactly the same.

Descriptive data of single participants mirror these effects. In VS 11 out of 12 participants (ca. 92%) showed better pointing performance in aligned trials (0°, ±90° and 180°) compared to oblique trials (±45°, ±135°). In ES only 3 out of 12 participants showed this pattern. 75% of the participants in this condition pointed more accurate from oblique body orientations.

As for latency, we tested employment of one vs. two reference axes. Fig. 5, right, depicts the indicative w- and m-contrast fits of the respective learning condition. Again, only positive m-contrast fits were of interest and compared to the corresponding w-contrast. In contrast to latency now positive values were produced for m-contrast centered on ±90° for VS condition (along short room axis) and m-contrast centered on 45°/−135° for ES condition (facing the wall of a corridor). The latter was larger than 0, $t(11) = 3.45$, $p = 0.005$, $d = 1.00$. However, no difference between m- and w-contrast fit was observed, ES: $t(11) = -1.75$, $p = 0.107$, $d_s = 0.51$, VS: $t(11) = 1.64$, $p = 0.130$, $d_s = 0.47$.

Further pointing results

Our results suggest that pointing performance varied as a function of corridor distance and body orientation and differently so for VS and ES learning. Did these factors cover most for the variability in the data or are there important communalities between the learning situations remaining, originating from the common object layout? For example, it might be easier to memorize and recall the position of the teapot compared to the position of the trumpet. In order to explore this, we calculated the residuals for latency and error that express unexplained variance after both, corridor distance as well as orientation during pointing, were accounted for. For each target location error and latency residuals were separately averaged over participants. We then calculated the correlation of these residuals between the two conditions. The same was done for each location participants were currently pointing from. Correlating the residuals of latency in VS and ES across the positions one is currently pointing from rendered a medium but non-significant result, n = 7, $r = 0.598$, $p = 0.156$. This correlation was mainly driven by the objects hammer and hair dryer. Both were located in the middle of the layout. Pointing from objects at the edge of the layout allows for a rather fast and rough estimation of the correct pointing direction since all remaining objects lie in somewhat similar direction. For example, being teleported to the trumpet, facing 0° (short wall in VS) any target object will lie behind you. In contrast, being positioned at hammer or hair dryer one is surrounded by targets. Pointing from there may naturally lead to longer decision times. We want to emphasize, however, that this correlation is not significant, thus, diminishing its importance. The remaining correlations for error residuals across current position and for error and latency residuals across target locations did not render significant results, $n = 7$, r's < 0.37, p's > 0.42. While some idiosyncrasies of the layout might be present in both conditions it seems that corridor distance and body orientation explain large parts of participants' performance.

Considering either latency or error in isolation can lead to huge misunderstandings of the data and effects reported above. Therefore, we looked into possible interrelations of latency and error in our data. Not a single participants showed a significant negative correlation between pointing error and latency, all $r > -0.16$, $p > 0.097$. Considering this analysis one can assume that our pointing data is not suffering from a speed-accuracy trade-off.

Order of object placement

We analyzed the order in which objects were relocated in the placement task. For each single participant we computed the Kendal-tau ordinal correlation as a measure of agreement between the relocation order in the placement task and the order of first encounter as predefined in ES learning (i.e., teapot, horse, hammer, banana, hair dryer, trumpet and racket). This correlation criterion reflects how much participants revert to this one possible learning order. The agreement of the placement order in the VS condition with ES learning order, indeed, is expected to be non-existent. It rather functions as an important baseline value ES correlation is compared against. Fig. 6, left and middle, visualizes the degree to which participants preserved the order of learning in ES and used it to relocate the objects in the placement task. When learning took place in ES a clear relationship emerged with an average $r = 0.82$. The correlation is larger than 0, $t(11) = 8.62$, $p < 0.001$, $d = 2.49$, and exceeds the correlation found in the VS condition, $r = 0.11$, $t(22) = -4.34$, $p < 0.001$, $d_s = 1.77$, which itself, as expected, does not differ from 0, $t(11) = 0.79$, $p = 0.446$, $d = 0.23$. The strong correlation found in ES is also supported by the individual data: 8 out of 12 participants provided a perfect match of $r = 1$, remaining participants uniformly showed a positive correlation. Hence, order induced by the learning process in the ES was still present in the context of conducting a configurational judgment task for which order was virtually irrelevant.

As VS memory might be bound to a specific order as well, we examined multiple potential alternatives. The examined orders were clustered in three groups: (1) Recall along rows and columns of the object layout (including 16 plausible orders), (2) following the random object presentation during the placement task (including eight plausible orders)[11] and (3) along corridor based

[11] Orders along rows/columns included, for example, from nearest to farthest row, each row from left to right; or from rightmost to leftmost column, across all columns alternating between nearest to farthest object and vice versa; et cetera. Orders following object presentation in the placement task when objects are arranged in random order along a single row included, for example, simply from left to right; or starting in the middle of the single row with the already active object to the leftmost object and from there the remaining object up until the rightmost object; et cetera.

encounter (single ES learning order as described above). The order in which the objects were presented to the participants in the placement task (i.e., a row) was randomized for each participant. However, orders along rows/columns of the uniform object layout are partly overlapping with the corridor order (mean correlation r = 0.50). For example, the potential order of learning and recalling the layout starting on the left and nearest object (teapot) and from there column by column to the right (each column starting with the nearest object) is highly associated with the corridor order, r = 0.81. A participant showing a high selection agreement with either will, thus, automatically show a high agreement with the other as well. Due to this non-independency a direct comparison between the strength of the different order correlations was not considered appropriate. Rather, we clustered participants with regard to the order that rendered the highest agreement with their individual placement order.

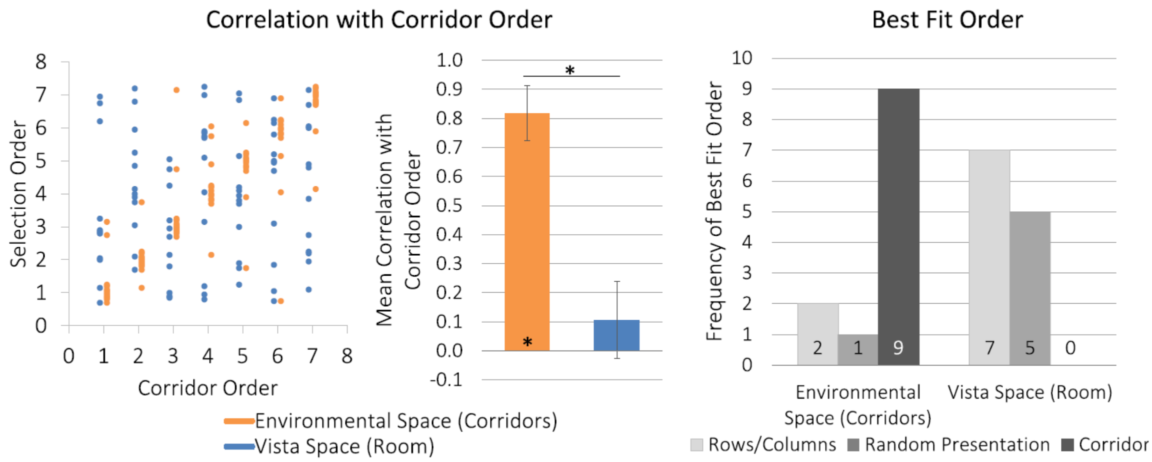


Fig. 6. Left: Placement order plotted as a function of the order objects were encountered when moving through the corridors. Middle: Mean correlation with corridor order. Right: Number of participants revealing best fit to either of three order clusters (along rows/columns, along object presentation during the placement task and along corridor based encounter) when rearranging the object layout from memory.

Fig. 6, right, depicts the amount of people scoring highest in any of the three order clusters.[12] The frequencies people were assigned to the three order groups differed significantly between the two learning environments, p = 0.001, two-tailed Fisher's exact test. More precisely, whereas in VS recall along rows/columns (7 participants; 3× best fit with order along rows, 4× best fit with

[12] For one participant of the VS condition the correlation of placement order with presentation order and row/column order was equally high. We considered presentation order as the more conservative, thus, more appropriate fit, as it does not assume a specific order preserved in memory.

order along columns) and along presentation of the objects in the placement task (5 participants) seems to be common, in ES most participants (9 out of 12) showed highest agreement with the corridor order.

Learning time and repetitions

Participants learning in ES needed significantly more time to meet the learning criterion (100% correct identification of objects at their corresponding location) compared to participants learning in VS, ES: M = 8.13 min, SD = 1.82, VS: M = 4.38 min, SD = 2.08, $t(22) = 4.71$, $p < 0.001$. This result is hardly surprising, since in ES movement was required to explore the environment. The number of repetitions needed to meet the learning criterion did not differ significantly between learning in ES, M = 1.25, SD = 0.13, and VS, M = 1.17, SD = 0.11, $t(22) = -0.48$, $p = 0.633$, $d_s = 0.19$.

Discussion

Many studies have examined spatial learning either in VS (e.g., Kelly and McNamara, 2008, Meilinger and Bülthoff, 2013, Mou and McNamara, 2002, Shelton and McNamara, 2001) or in ES (e.g., Avraamides and Kelly, 2010, Brockmole and Wang, 2002, Brockmole and Wang, 2003, Cohen et al., 1978, Kelly et al., 2007, Kosslyn et al., 1974, McNamara, 1986, Meilinger et al., 2014, Newcombe and Liben, 1982, Wang and Brockmole, 2003a, Wang and Brockmole, 2003b, Werner and Schmidt, 1999). The present experiment aimed at answering the question of how spatial memory is different for VS and ES by comparing memory for exactly the same object layout. In summary, the observed distance, order, and alignment effects indicate that spatial memory for ES and VS differ in terms of preserved spatio-temporal learning experience, and employed reference frame orientation.

We examined whether the process of estimating object-to-object relations depended on corridor or Euclidean distance to the target. To date, better performance for closer targets with regards to traveled distances in ES was mainly found in the inspection of errors (Thorndyke & Hayes-Roth, 1982), but not latency. However, since an increase in error over corridor distance could result also from error accumulation during learning (thus, memory precision), we assume latency to be a more suitable criterion for indicating processing time (thus, retrieval) (see also Pantelides et al., 2016). We hypothesized an incremental process when recalling memory acquired in ES. Indeed, when learning took place in ES results from the visual pointing task suggest that pointing latency increased with increasing corridor distance. With each additional corridor separating participant and target, more time was needed to point to the target. To our knowledge, this was the first time a traveled distance effect on retrieval latency of survey memory has been demonstrated. It points to the fact that in ES the process of recalling locations beyond the current VS cannot be a simple and direct read-out from a single memory unit. Estimation of object-to-object relations in ES rather happens incrementally during pointing, successively activating spatially and/or temporally related information from memory. Whether this process consists of mentally walking down the route (Byrne et al., 2007, Sanders et al., 2015) or of constructing a mental model of the non-visible environment parts from one's current location (Meilinger, 2008), however, remains an issue to exam further.

The layout itself was not responsible for the corridor distance effect, since the effect was not evoked by VS learning. Similarly, pointing latency did not increase with increasing Euclidean distance between object pairs. As McNamara (1986) showed that Euclidean distance affects the accuracy of survey judgements, these results again support the disparity of error and latency measurements. Assuming that pointing latency is a manifestation of the recall process rather than the precision of memory, all objects learned within the VS seemed to be retrieved equally fast irrespective of both distance metrics. Likely, no further mental constructions were needed here; rather, configurational memory already comprised all object locations in a single reference frame as suggested by prior research (McNamara et al., 2008, Mou et al., 2004). Additionally, this result might indicate that the positive association between pointing latency and Euclidean distance in ES was merely driven by the correlation of corridor and Euclidean distance. Taken together, learning differs across ES and VS environments which manifests in the absence of a distance effect in the VS environment, and an effect of corridor distance in the ES environment. Specifically, even though both groups of participants were instructed to learn the same object layout, in the ES environment the temporal and spatial learning procedure was preserved in memory retrieval, and thus, the structure of the layout representation.

We predicted that retrieving memory for ES would be bound to the learning order due to the successive entering of corridors in ES learning. Indeed, a high correlation between the order of first object encounter in ES and the order of object relocation during the placement task was found for ES. It is important to note that this order effect was observed in the retrieval procedure of a survey task, a task that did not require placing the objects along a specific order. In fact, in order to select objects according to the encoding order in ES learning, participants even had to deselect items, disregard the random objects order presented in front of them and deliberately select other items. The strong relation between encoding order and placement order in ES is in contradiction to a VS process of reading-out from a single reference frame representation (McNamara et al., 2008, Mou et al., 2004), which itself does not predict a preferred order, as mere inter-object relations are memorized. Interestingly, from twelve participants in the ES condition, eight provided a perfect order match and all others showed a positive correlation as well. Hence, never was an object layout reconstructed from the endpoint of the ES to the starting position, even though this sequence was encountered just as often as walking from start to end. Previous research showed how explicitly landmark sequence is preserved and affects memory retrieval when learning a route (Janzen, 2006, Schweizer et al., 1998, Strickrodt et al., 2015, Wiener et al., 2012) and when learning the configuration of an ES (Moar & Carleton, 1982). The present experiment extends these findings by demonstrating that order effects are determined by the very first experience within the environment (original forwards direction) even when walking in both directions (forwards and return path).

Similar to the effect of corridor distance, we did not expect the learning order of ES to be reflected in spatial memory acquired in a VS. VS learning functioned as a control condition and the absence of a meaningful correlation with corridor order demonstrates that effects in ES were not due to the layout structure itself. Regarding the best order fit analysis, placement order was much less consistent in VS. Participants showed placement order patterns along rows and columns (with no systematic preference for either following rows or columns) or merely along the random presentation order in the placement task. From this result we cannot determine whether layout memory of VS is generally structured along rows and columns but easily overwritten by presenting alternative sequences or whether just some participant preserve a row/columns structure in memory. It should be noted that we consider a statistical comparison within a learning condition as problematic as the amount of potential orders that were covered in our analysis differed between rows/columns, presentation and corridor order. We examined 16 plausible orders representing selection along rows and columns, but only eight plausible orders of following the random object presentation in the placement task and one order representing corridor learning. Thus, the chance to be clustered in either order group is unequal. This is problematic for the distributions of orders within a learning group, but less so for comparisons between learning groups. It should also be noted that the analysis of a best fit sequence depends on the selected orders that are taken into account. We clearly did not cover all feasible orders, but only a subset of reasonable orders. As a result, most participants have been clustered into order groups because all remaining correlation rendered lower values – even though their maximum correlation was not significant. In sum, we showed a clear difference between VS and ES learning. ES learning clearly preserved the distinctive initial learning encounter.

As expected, we found evidence for differently oriented reference frames in VS and ES indicated both by error and latency. W-contrast fits were positive for learning in VS, hence, reference frames were centered on the aligned orientation of initial view, room geometry and object layout (0°, ±90°, 180°) as in prior studies (e.g., Kelly and McNamara, 2008, Shelton and McNamara, 2001, Valiquette and McNamara, 2007). In contrast, in ES negative w-contrasts were found, i.e., overall lowest error rate and fastest responses were shown when aligned with or orthogonal to the corridors (±45° and ±135°). Such an alignment with vista units of an ES has also been observed before (Meilinger et al., 2014, Werner and Schmidt, 1999). Our study demonstrated in addition that the initial view and the global layout-intrinsic orientation are less important for setting the reference frame in an ES. In VS studies it has already been shown that the initial view can be dominated by another, if this new view is aligned with a geometric feature (e.g., global room, mat, object layout) (Kelly and McNamara, 2008, Shelton and McNamara, 2001, Valiquette and McNamara, 2007). Similarly, the same seems to account for ES learning. During walking participants are aligned with the corridor walls as well as with the locally visible objects. These factors seem to determine the alignment of spatial memory.

We further examined whether the assumption of two orthogonal axes constituting a reference frame, i.e., performance following a w-shaped pattern, holds. Alternatively, space could be represented along a single axis, i.e., performance following an m-shaped pattern. Considering both, pointing error and latency, our results are inconclusive about whether a single or two orthogonally aligned reference axes were involved. Pointing latency was better described by an m-contrast centered on the long axis of the room (0°/180°) in VS, and by an m-contrast centered on the corridor axis (−45/°135°) in ES compared to the corresponding w-contrasts. In contrast, when analyzing pointing error in ES the m-contrast centered on the orientations when facing a corridor wall (45°/−135°) produced a positive fit. Also, VS now evidenced two reference axes rather than one. From this no clear conclusions about the number of reference frame axes can be made. Nonetheless, our results demonstrated that performance pattern in ES are in clear opposition to the performance pattern in VS learning, which we assume to be the crucial point here. The overall orientation of the reference frame seems to be well captured in the w-contrast fit rendering this measurement a reliable, even rather conservative mean to detect differences in reference frame alignment between ES and VS.

In addition to the alignment effect, VS learning also resulted in higher pointing accuracy than ES learning. This difference likely originated from the specific differences of VS and ES learning such as successive vs. instant visibility of objects, required movement and common visible anchor (i.e., the room) for VS, but not ES learning. In ES, participants need to relate locations that were never encountered together. The mental effort to construct a mental representation of the object layout is likely to be higher and the process more error-prone in ES compared to VS learning.

This is the first work showing that retrieving configurational memory for ES is bound to the traveled distance and order of learning. Observed effects cannot be accounted for by a simple read-out process from a single reference frame, which typically explains memory retrieval for VS. Albeit these results clearly show where memory for VS and ES differ, they do not answer the question of the underlying reasons, which is the subject of Experiment 2.

Experiment 2

Experiment 2 was concerned with what aspects of the learning situations may cause differences in the memory structure. Most importantly, the separation of ES in multiple VS units (compartmentalization), the movement through space, and the successive encounter of objects should be treated as potentially relevant factors for a divergence.

Compartmentalization

The nature of ES is that the environment is separated into units by spatial borders. Opaque barriers were found to elicit overestimation of physical distance between targets (Kosslyn et al., 1974). The effects of distance, order, and reference frame alignment found in Experiment 1 identify additional characteristics on which ES memory differs from VS memory which, indeed, might have been caused by opaque borders. There is, however, evidence that not just opaque borders, but also non-opaque borders elicit distinct distortions in spatial judgements. McNamara (1986), for example, reported a bias in distance estimation when learning an environment in which spatial borders were merely set by strings on the floor (i.e., no opaque border). As the compartmentalization of space in the ES condition was inevitably linked to the need of movement and to successive object encounter it is important to identify the cause of effects found in Experiment 1. In the VS condition of Experiment 1 participants were restricted to learn the fully visible object layout standing at one location. To account for the potential influence of movement and successive object presentation we emulated both in a VS setting in Experiment 2. Indeed, in a real-world scenario one can easily move around in VS as well, successively passing the objects within. However, exploring a VS is not subject to restrictions comparable to restrictions imposed by an ES structure. Most importantly, a VS provides a common reference space objects within the space can jointly be related to. This common reference space might facilitate the integration of object locations into a single reference frame compared to learning in a compartmentalized space (ES), regardless of movement and successive objects presentation.

Movement

The translation through space when learning an ES makes it possible to experience a multitude of visual and proprioceptive information. In contrast, many studies concerned with the learning of object layouts in VS typically exclude walking from the learning procedure. Often, visual information are presented from one up to a few predefined vantage points. Indeed, learning in VS does per definition not require movement, since it involves all spatial information that can be gathered from a single vantage point.

To examine whether the effects of order, distance and alignment originate from movement, in Experiment 2 we now had participants walk through the room along a path matching exactly the movement through ES. Now the path determined, for example, that in order to travel from the teapot to the hammer the horse has to be passed by making a detour. The prevention of a direct path between teapot and hammer might be interpreted as a boundary, which in turn might influence memory structure. Furthermore, now a walking distance between pairs of objects was provided. Both spatio-temporal information, the impression of a non-visible boundary and the experienced walking distance, might promote order and distance effects. At the same time, these effects might be diminished since the VS itself allows for a global observation of all environmental features.

Finally, walking across the object layout might also induce a different reference frame orientation. By introducing the path traveled in ES in a single room we created a conflict between multiple inputs. The visible context of the room (room geometry, global object layout) and the initial view now have to compete against varying viewpoints, perspectives and body orientations during movement with the main learning orientation being oblique to the room axes.

Previous papers have demonstrated the importance of self-to-geometry alignment that is experienced later during learning, after the initial view for setting the reference frame orientation (Kelly & McNamara, 2008). Also bodily cues were found to be of importance. Yamamoto and Shelton, 2005, Yamamoto and Shelton, 2007 showed that proprioceptive learning (blindfolded walking) by itself can yield a reference frame orientation seemingly independent of and comparable in strength to visual learning. Hence, multiple encoded views and different body orientations during learning might counterbalance the visible context and influence reference frame usage. Varying whether participants walk along the route or exhibit the object layout from a constant view will help to understand how these factors influence the structure of spatial memory.

Object presentation

Another aspect distinguishing learning in ES from learning in VS is that the visibility of objects is not simultaneous. The environmental borders and transition points from one spatial entity to the next determine the sequence in which objects are encountered. While objects of the previous VS will be out of sight, objects in the currently visited VS will now be attended. As an important aspect of the learning procedure in ES, we wanted to examine whether successive presentation cause or contribute to the maintenance of spatio-temporal encoding information in memory, i.e., distance and order effects, and to the alignment of the reference frame. Therefore, in Experiment 2 some of the participants were confronted with the target objects step-by-step, adopting the object encounter of the ES condition of Experiment 1 within a VS. Objects located within the same corridor in the ES condition of Experiment 1 were now, within the VS room, visible at the same time, alternating with the next object pair and so forth. Such a learning procedure will set the spotlight to discrete object pairs while preventing the view upon another proportion of the layout objects. Thus, similar to movement through space, successive object presentation determines a specific spatio-temporal learning experience that might as well induce order and distance effects. Furthermore, now the global object layout as a potential cue influencing reference frame alignment, will not be apparent anymore. Rather, pairs of objects aligned with orientations oblique to the room geometry constitute another visual cue, which might affect reference frame alignment.

In Experiment 2 we set out to examine which aspects of the learning procedure that distinguish ES from VS learning lead to divergence in the spatial representation of the same object layout. As spatial separation along opaque barriers cannot be varied independently of movement and successive object presentation we eliminate the compartmentalization of space, but varied the other two factors. Learning conditions in VS were step-by-step adapted to ES learning. We had three learning groups: Participants viewed objects successively from a static position (stat-succ), viewed the objects simultaneously but followed the path executed in ES (move-simult), or viewed the objects successively while following the path (move-succ). The last condition differed from ES learning only by the absence of walls, i.e., the absence of multiple VS units. In combination with VS learning in Experiment 1 (stat-simult) this yielded a 2 (object presentation: simultaneous vs. successive) × 2 (movement: static position vs. movement through space) plan to examine how far any of these conditions leads to results matching findings of ES. Finding that movement or successive object presentation (or their combination) in VS elicit similar effects to learning in ES would assign them to be determining factors for configurational learning of ES. However, the absence of order effects, distance effects, or reference frame alignment along oblique orientations in Experiment 2 would identify the separation of space as the remaining, determining factor for the distinct construction of configurational knowledge.

Method

Methods were identical to Experiment 1 except for the alternations described.

Participants

36 participants (19 females) with a mean age of 26.97 years participated (*SD* = 7.57, [16;48]) and were randomly assigned to one of the three groups (12 per group). One participant withdrew from participation after completing the visual pointing task; hence, placement data of this participant was not recorded. From the original sample of 39, two participants were excluded since they did not perform significantly better than chance level of 90° absolute pointing error. Another participant was excluded due to computer problems during the pointing task.

Materials and procedure

All participants learned the object layout within the rectangular room of the VS condition of Experiment 1 and the same initial view (Fig. 1, right). The remaining procedure was adjusted according to the conditions. In condition stat-succ no movement was required. Participants were not allowed to leave their current position, but they were obliged to look around. Objects were presented in successive order matching the presentation of objects of ES learning. Objects formerly presented within the same corridor in ES were now presented at the same time (object pairs). Common visibility of objects, henceforth, was as follows: teapot – horse and hammer – banana and hair dryer – trumpet and racket. To enable self-paced learning participants pressed the button of a controller. Following a duration of continued button press a switch from one object (pair) to the next took place. This duration matched average walking time through a corridor in ES as determined in pre-experiments. Importantly, when the former object (pair) disappeared, the next object (pair) appeared. Since in ES learning of Experiment 1 participants were allowed to stop at any point during their movement through space and, hence, determined encoding time themselves, also participants in the stat-succ condition could prolong the view upon the current object (pair) by pausing the button press of the controller. When reaching the last object pair (trumpet and racket), object pairs were presented in backwards order again. This procedure was repeated one more time (similar to walking twice from start to the end point in ES).

Participants in the move-simult condition moved through VS, matching the path through the ES corridors of Experiment 1. Grey discs on the floor led participants on a specified path. When reaching a disc, the next target disc lit up. Disc locations corresponded to the location of direction change in ES, i.e., the end and start point of each corridor. Only the disc at the current and the next position was displayed. Participants had to walk on a straight line towards the next disc without detours. However, they were allowed to stop at any point and to look around. Participants in the movement conditions had to follow the corridor route four times, alternating between forward and return path. In the move-simult condition the complete object layout was visible simultaneously during learning, analogous to VS learning in Experiment 1.

The move-succ condition matched both the movement and the visibility of objects to ES learning. Objects were presented successively, as in the stat-succ condition, while participants had to follow the path mirroring movement in ES similar to the move-simult condition. The presentation of object pairs occurred automatically when participants reached positions that corresponded to the passage between two ES corridors.

After translation between start and endpoint (movement conditions) or being confronted with all object pairs four times (stat-succ), the objects were removed and acquired object knowledge was tested. For the static learning condition, the learning test was identical to VS learning in Experiment 1. For the movement conditions, the learning test was identical to the environmental condition of Experiment 1, except that participants walked through the room, not the corridors and their walking was again guided by the discs. After reaching the learning criterion (100% correct identifications) the test phase started.

Data analysis

From the 36 participants, ca. 4% of the pointing performance data was deleted due to deviation of more than 2 *SD* from a participant's overall mean. One participant in the move-simult condition withdrew from continuing the experiment after the pointing task, rendering only eleven participants in this condition for the analysis of placement order. The aim of Experiment 2 was to understand the impact of movement, successive object presentation and compartmentalization of space on the acquired memory. We focused on quantitative effects on which VS and ES condition clearly differed in Experiment 1. Thus, we confined our analysis to latency increase with movement/presentation distance (i.e., corridor distance) as represented by individual regression slopes, pointing performance fits to a w-pattern centered on 0°/±90°/180° and correlations of placement order with ES learning order (i.e., corridor order). These analyses enabled us to judge how much movement, object presentation and their combination render equal values to those obtained from ES learning. We analyzed these data with a 2 (movement) × 2 (object presentation) ANOVA (including VS condition of Experiment 1) to estimate the influence of each factor separately, as well as of their combination. Subsequently, we compared the level of each parameter with the respective value in the ES condition of Experiment 1. Please note that comparisons across experiments were valid as Experiment 1 and 2 were conducted together and participants were assigned randomly to all conditions.

Results

Distance to target

For the distance analysis of Experiment 2 we regressed pointing latency of each single participant onto the two distance metrics and extracted the standardized slopes b. Fig. 7 depicts the mean slopes for the linear increase in pointing latency as a function of corridor distance (left panel) or Euclidean distance (right panel) to the target. The three bars on the right in each panel mark data collected in Experiment 2. Data from Experiment 1 is depicted again in the two bars on the left for ease of comparison. Note that corridor distance in Experiment 2 was not induced by actual corridors but by walking and successive layout experience. The ANOVA on the linear increase of latency across ascending corridor distances revealed that neither movement nor object presentation exert a main effect on the level of slope. Further, no significant interaction could be found, *F's* < 2.1, *p's* > 0.159. Neither learning condition led to a slope larger than 0, *t's* < 0.95, *p's* > 0.363, similar to results of the VS condition in Experiment 1 (stat-simult). Thus, the manipulation of movement and object presentation in VS did not lead to an increase of pointing latency across ascending corridor distances. When considering individual slopes, in each new VS condition 6 to 7 out of 12 participants revealed a positive slope. This further supports the aforementioned results. Consistently, all VS slopes differed from the slope attained in ES, *t's* > 4.01, *p's* < 0.002. Since neither movement, nor object presentation or the interaction rendered effects this suggests that the increase in reaction time across corridor distance found in Experiment 1 was due to the visual separation of the space.

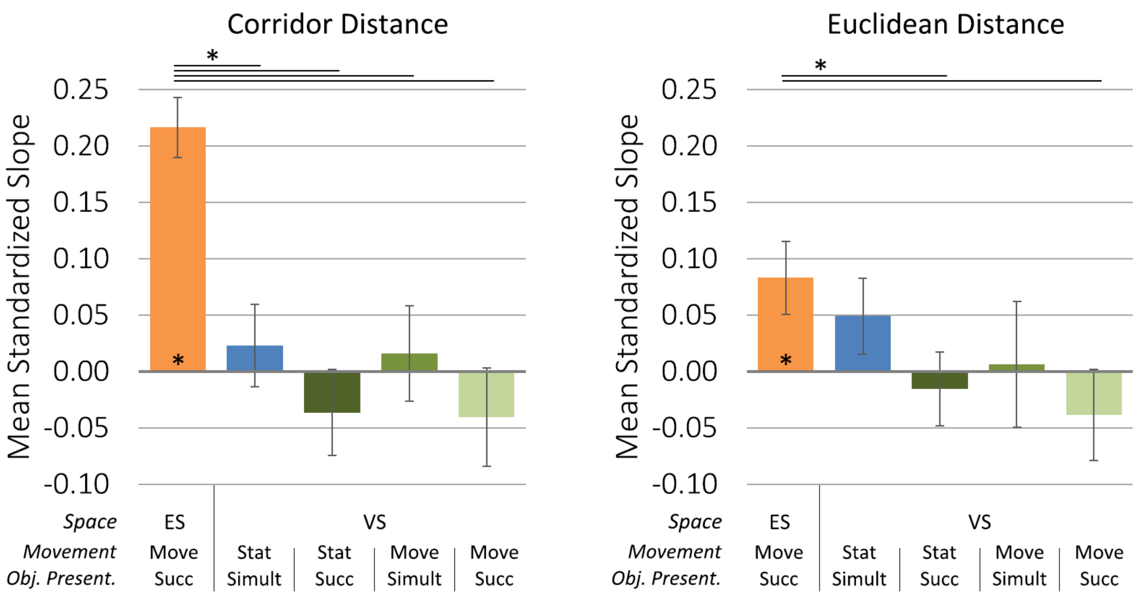


Fig. 7. Mean standardized slopes of the effect of corridor distance (left) and Euclidean distance (right) on pointing latency. The three bars on the right of each panel represent the new learning conditions of Experiment 2, mirroring ES learning in a single room. Move = Movement, Stat = stationary without movement, Succ = successive presentation of objects, Simult = simultaneous presentation of objects.

A similar analysis was conducted for the effect of Euclidean distance. Standardized slopes representing the potential linear increase of latency across Euclidean distance were submitted to an ANOVA. Results mainly resembled those attained for corridor distance: No main effects or interaction of movement and object presentation could be found, *F's* < 1.71, *p's* > 0.197. Also, neither VS condition that mirrored aspects of ES learning rendered slopes that exceed 0, *t's* < 0.96, *p's* > 0.359. Individual slope distribution was again near chance: 5 of 12 participants (ca. 42%) in each new VS condition yielded a positive slope. When comparing the linear increase of latency across ascending Euclidean distance obtained in this experiment to the ES condition from Experiment 1, only conditions stat-succ and move-succ differed significantly from ES, *t's* > 2.13, *p's* < 0.045. This supports results found in Experiment 1. Euclidean distance again was of no significant importance when learning took place in a single room (VS).

Reference frame orientation

Pointing latency

Pointing latency as a function of body orientation is depicted in Fig. 8, top left. Contrast fits to a w-pattern centered on 0/±90°/180° are displayed on the top right, the three bars on the right of the panel marking data collected in Experiment 2. No main effects of movement or object presentation on w-contrast fit were found, *F's* < 0.17, *p's* > 0.676. There was, however, a trend for an interaction of movement * object presentation, $F(1, 44) = 3.256$, $p = 0.078$, $\eta_p^2 = 0.069$. Condition stat-succ rendered highest contrast fits. Average contrast fits for pointing latency in all VS conditions were positive, for stat-succ significantly above 0, $t(11) = 2.37$, $p = 0.037$, $d = 0.68$. In line with the results of Experiment 1, fits were clearly different from ES learning, *t's* > 2.30, *p's* < 0.032. Thus, neither movement nor successive presentation (or their combination) yielded a similar shift in reference frame orientation towards oblique directions (±45°, ±135°) as found in ES learning. The same pattern was found when looking at single participants: In the stat-succ, move-simult and move-succ condition 9 (75%), 7 (58%) and 6 (50%) out of 12 participants showed numerically faster pointing performance in trials of aligned body orientation (0°/±90°/180°) compared to trials of oblique body orientation (±45°/±135°), respectively.

Absolute pointing error

Fig. 8, bottom, displays results for the absolute pointing error which parallel those of pointing latency. No main effect of movement or object presentation or an interaction between the two could be found, *F's* < 0.51, *p's* > 0.482. Again, presenting objects successively or allowing translation through space, did not seem to influence w-contrast fits, i.e., the selection of reference frame orientation. Contrast fit for move-simult exceeded 0, $t(11) = 2.96$, $p = 0.013$, $d = 0.85$, and for stat-succ by trend also, $t(11) = 2.08$, $p = 0.062$, $d = 0.60$. Just as for pointing latency, w-fits in all VS conditions were numerically larger than 0 and significantly different from the contrast fit of ES learning of Experiment 1, *t's* > 2.10, *p's* < 0.048. This pattern again was mirrored in individual data: In the stat-succ, move-simult and move-succ condition 8 (67%), 10 (83%) and 9 (75%) out of 12 participants pointed more accurate in trials of room aligned body orientation (0°/±90°/180°) compared to trials of oblique body orientation, respectively.

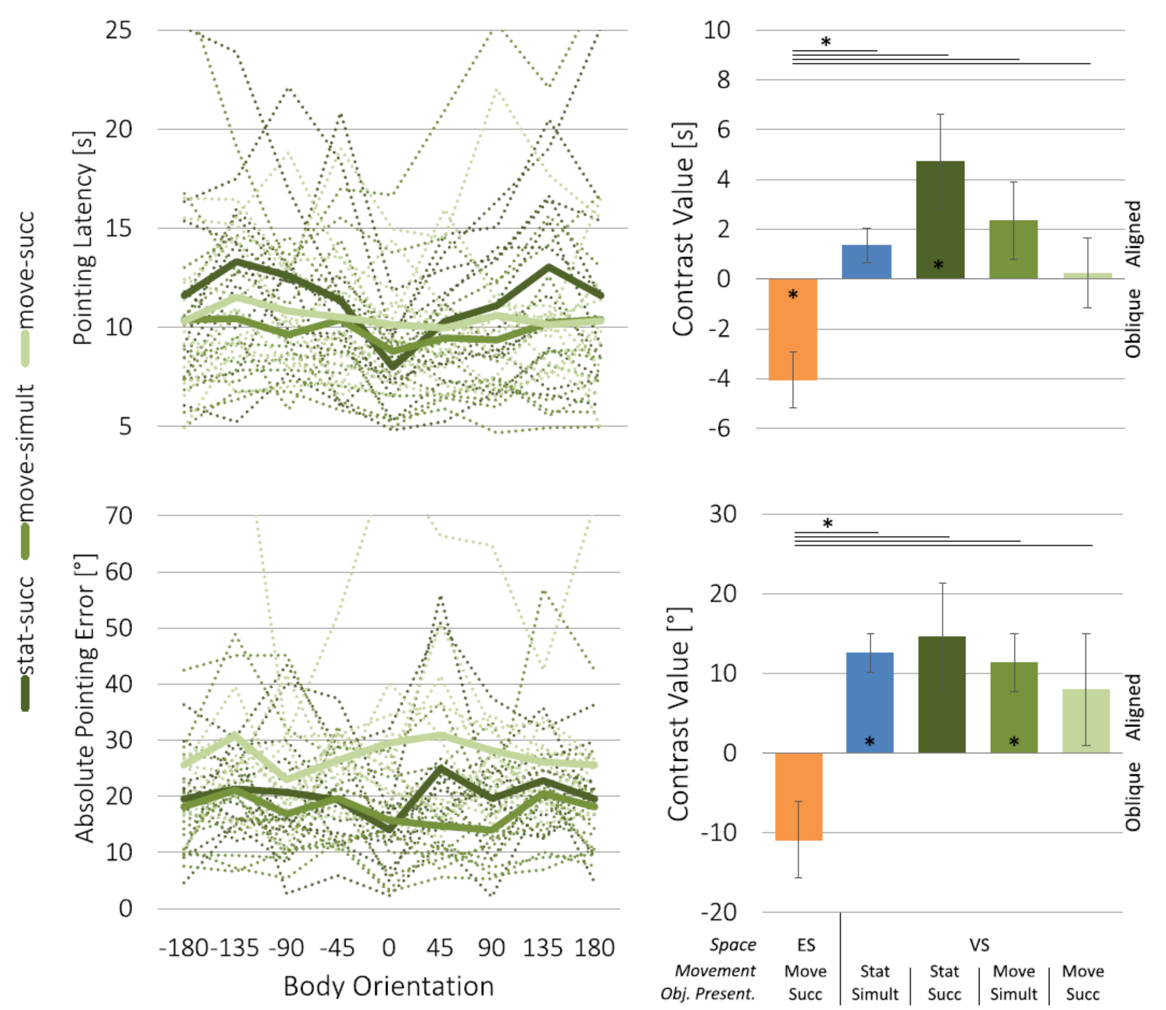


Fig. 8. Pointing latency (top) and absolute pointing error (bottom) as a function of body orientation. Left: Individual (dashed line) and averaged (solid line) pointing performance of the new VS learning conditions of Experiment 2. Right: Data fit to the w-contrast centered on 0°/±90°/180°.

Pointing latency and error were negatively correlated in 1 out of 36 participants, $r = -0.31$, $p = 0.006$, remaining correlations $r < 0.21$. The average correlation of error and latency across the three VS condition of Experiment 2 was $r = 0.01$. As in Experiment 1, data does not seem to exhibit a speed-accuracy trade-off.

Order of object placement

Fig. 9 shows the mean correlations between experienced order and placement order. Considering the different VS conditions neither a main effect of movement or object presentation, nor a significant interaction was present, *F's* < 2.14, *p's* > 0.151. Thus, varying movement and object presentation in VS did not seem to have influenced the corridor order effect. However, in contrast to VS from Experiment 1 the mean correlations of stat-succ, $t(11) = 3.30$, $p = 0.007$, $d = 0.95$, and move-simult, $t(11) = 3.49$, $p = 0.006$, $d = 1.05$, differed from

0, and by trend also move-succ, $t(11) = 1.861$, $p = 0.090$, $d = 0.54$. Although lager than 0, the correlations were still smaller than in ES for move-simult, $t(21) = -2.42$, $p = 0.024$, $d_s = 1.02$, move-succ, $t(16.1) = -2.15$, $p = 0.047$, $d_s = 0.88$, and by trend also stat-succ, $t(22) = -2.05$, $p = 0.053$, $d_s = 0.84$. Consequently, it can be inferred that movement and successive object presentation either alone or in combination led to a mediocre relocation preference along the order of learning. However, the order effect still differed from the effect found in ES learning. When controlling for gender as a covariate, a main effect of gender was found, $F(1, 44) = 6.583$, $p = 0.014$, $\eta_p^2 = 0.141$, which, however, did not yield any major changes in abovementioned results. Overall, female participants exhibited a larger correlation between learning and placing order, $r = 0.54$, than males, $r = 0.13$, but both seem to be similarly affected by movement and object presentation.

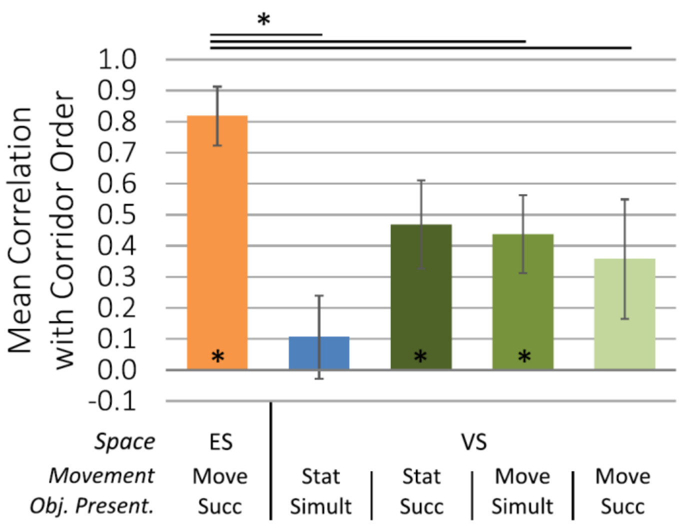


Fig. 9. Mean correlations of corridor order and placement order for each learning condition.

Learning time and repetitions

The time needed to learn the environment varied across the different VS conditions. We observed a significant main effect of movement, $F(1, 44) = 19.06$, $p < 0.001$, $\eta_p^2 = 0.30$, as well as a trend for object presentation, $F(1, 44) = 3.47$, $p = 0.069$, $\eta_p^2 = 0.07$, on learning time. Learning took more time when movement and successive object presentation was introduced. The interaction of movement and object presentation was not significant, $F(1, 44) = 0.29$, $p = 0.598$, $\eta_p^2 = 0.006$. Participants spent on average 6.77 min ($SD = 3.20$) learning in the stat-succ condition, 11.26 min ($SD = 4.28$) in the move-simult condition, and 15.55 min ($SD = 11.02$) in the move-succ condition.

Average number of learning repetitions required to pass the learning criterion (100% correct identification of objects at their corresponding location) were *M* = 1.42, *SD* = 0.15 for stat-succ, *M* = 1.75, *SD* = 0.18 for move-simult and *M* = 1.83, *SD* = 0.27 for move-succ. Movement conditions evoked more learning repetitions than learning from a static position, as indicated by the main effect of movement, $F(1, 44) = 7.14$, $p = 0.011$, $\eta_p^2 = 0.14$. No main effect of object presentation or interaction of object presentation × movement could be found, *F's* < 0.80, *p's* > 0.377. Findings suggest that movement makes it harder to learn the locations of objects in space. Nevertheless, individual adaptation of learning time and number of repetitions ensured that acquired object location knowledge was sufficiently comparable between groups.

Discussion

Experiment 2 investigated possible underlying mechanisms differentiating VS and ES learning. Learning experience in ES differed from VS learning. Specifically, in ES navigators were confronted with a compartmentalized space, had to walk through the environment and experienced successive presentation of objects. This study manipulated movement and successive object presentation and their combination within the visual context of a VS to estimate their respective and combined influence. Our results across all three parameters revealed a clear picture: neither changes in movement, object presentation or their combination influenced the acquired layout memory in a way equal to learning in ES. In fact, basically each single VS condition differed on each parameter from the ES condition (only learning order for stat-succ did – with $p = 0.053$ – not reach significance). We conclude that the separation via opaque borders must be the main source of differentiation.

No evidence could be found, that either the successive visibility of objects, movement through space or the combination of both had a specific effect on pointing latency with increasing movement/presentation distance (i.e., corridor distance). In Experiment 2 no additional time was needed to activate memory for objects which were passed and/or perceived later during learning, regardless of the strong spatio-temporal characteristics of movement and successive object presentation. Movement, providing additional proprioceptive input, has previously been found to shape spatial knowledge (Chrastil and Warren, 2013, Waller et al., 2004, Yamamoto and Shelton, 2005, Yamamoto and Shelton, 2007). However, bodily walking cues did not yield an ES-like memory structure when learning in VS. Similar pointing latencies for both corridor and Euclidean distance also strengthen the conclusion that object locations in each VS condition were memorized within a single, integrated representation, irrespective of the learning procedure and straight-line distances between objects. This implies that the presence of borders between corridors in ES learning is responsible for the incremental processing during retrieval that was found in Experiment 1.

In the placement task we observed medium size order effects. The correlation between learning order and placement order were larger than 0 in every VS condition incorporating movement and/or successive object presentation (for move-succ at least by trend). This dissociates VS conditions of Experiment 2 from the original VS condition of Experiment 1 (stat-simult), where no order effect was present. Creating a spatio-temporal contingency by guiding movement and/or restricting the attentional focus led to the incorporation of learning order in configurational memory for VS as well, guiding memory retrieval, however, to a lesser extent compared to ES learning. Sensitivity for route direction or order was mainly shown in studies utilizing ES (e.g., Janzen, 2006, Moar and Carleton, 1982, Schweizer et al., 1998, Wiener et al., 2012). Our results imply that this sensitivity seems to be – at least partly – independent of whether there is a common reference space (VS) or not (ES). Most importantly, none of the VS conditions induced similar learning order effects as ES learning. This illustrates that guidance of attention can only partly explain the order effect found in ES and it reveals the impact of opaque borders on shaping order dependency. Relying on order when learning in ES or uncoupling from the learning order in VS can each for itself yield advantages. In ES order might be particularly advantageous to not confuse the sequence of single corridors. This is exceedingly important if ES memory consists of multiple subunits. In contrast, VS memory that is not bound to a specific order might be retrieved more flexibly.

Why did movement and successive object presentation in VS yield middle sized order effects, but no effects of movement/presentation distance (i.e., corridor distance)? We speculate that placement order in the layout reproduction task (irrespective of where exactly participants place the objects) is associated with the temporal aspect of spatial knowledge, whereas pointing latency across varying distances captures how spatial aspects (direct relations between pairs of objects) are retrieved from memory. Such a dissociation of memory systems specialized in spatial locations vs. behavioral responses, which incorporate also the temporal order, have been proposed before (Packard and McGaugh, 1996, Restle, 1957). In a case study van der Ham et al. (2010) demonstrated how temporal and spatial aspects of navigation are dissociated in humans. Impairment in a route ordering task did not similarly lead to impairment in route continuation task, or vice versa. Likewise, addressing different aspects of survey knowledge might be prone to an analog dissociation between temporal and spatial aspects. Furthermore, predefining an order by movement or successive object presentation might generate an additional verbal memory trace constructed along the learning order. Verbal memory was shown to be involved within route learning (Meilinger et al., 2008, Wen et al., 2011) as well as learning of an object layout (Meilinger & Bülthoff, 2013). Memory retrieval in the subsequent placement task might be initialized following this verbal code. Female participants exhibited larger order correlations than males. This was the only effect of gender observed in both experiments. We speculate that this effect might originate from a stronger reliance on a verbal coding strategy for spatial material in women (Coluccia & Louse, 2004). Thus, the dissociation between spatial and temporal aspects of spatial memory may explain the emergence of mediocre order effects in the absence of distance effects, and/or verbal coding along the learning experience might be responsible for part of the order effect observed.

In Experiment 2, we induced a conflict of available reference axes evoked by movement, successive object presentation and a common reference space. Participants moved along paths and/or were confronted with pairs of objects which were aligned with an axis (main axis of −45° to 135°) that is oblique to the initial view, room geometry and global object layout (0°/±90°/180°). W-contrasts of all VS conditions differed clearly from ES learning. This does not imply that movement and successive object presentation have no effect on the alignment of the reference frame. For example, the pattern of orientation dependency of the pointing performance in move-succ (most similar to ES learning) seems to become more leveled, not showing a clear trend in either direction. Here the maximum conflict of available geometric axes and views is experienced. However, as the pattern even in this condition induced no conversion of the dependency pattern of body orientation and clearly differed from ES learning, we conclude that on their own movement and successive object presentation are not sufficient to assimilate the clear reference frame alignment along oblique orientations that was found in ES. Following this, we conclude that compartmentalization through opaque barriers – the remaining factor differentiating ES and VS learning of Experiment 1 – was responsible for the clear shift to oblique orientations in ES. Within a single corridor of the ES the visible objects, corridor walls and experienced views through movement were jointly aligned, supporting a corresponding reference frame alignment (Kelly and McNamara, 2008, Shelton and McNamara, 2001, Valiquette and

McNamara, 2007). We assume that the potentially conflicting cue of the initial view was easily overwritten by the viewer-space-alignment when walking through the corridors (compare to Kelly & McNamara, 2008). Furthermore, the opaque borders literally cut off the perception of the potential conflicting cue of the global layout orientation. This could only be inferred at the moment the last object was discovered and indeed only a mental, probably distorted representation could have been used. As our results demonstrated, no effortful restructuring and realignment of layout memory on the basis of an inferred global layout orientation – a cue that extends beyond the current corridor unit – was carried out in ES. Both in VS and ES the visible surrounding geometry seems to serve as the main cue organizing a reference frame for remembering locations in space.

In conclusion, our results clearly show that movement and object presentation introduced in a VS do not render the performance pattern observed in ES learning. This leaves the compartmentalization of space as a main factor causing the memory structure of an ES to differ from the memory structure of a VS. Having a common, continuously visible reference within the VS allowed participants to (1) integrate all target locations into one representation without successively activating spatially distant information from memory, (2) rely less on the order of learning although full decoupling was not observed, and (3) to employ reference frames different from the ones used in ES learning.

General discussion

We examined memory for an object configuration learned within a VS (a single room) or within an ES that is spread across multiple corridors. Experiment 1 showed that configurational memory differed qualitatively: Contrary to VS learning, retrieving memory of the ES was bound to the distance experienced and to the order in which the objects were learned. Also, ES learning employed different reference frames whose orientation followed the orientation of corridors rather than the initial view of the environment or the layout-intrinsic orientation. Experiment 2 revealed that neither the movement trajectory, nor the successive presentation of objects, or the combination of both could fully account for the qualitative differences. Having examined these factors we conclude that compartmentalization into multiple sub-spaces is the main factor responsible for the dissociation of memory between these two classes of space.

Our results blend nicely into existing findings. Spatial borders were found to affect updating (Avraamides and Kelly, 2010, Kelly et al., 2007, Wang and Brockmole, 2003a, Wang and Brockmole, 2003b), distance estimation (Cohen et al., 1978, Kosslyn et al., 1974, McNamara, 1986, Newcombe and Liben, 1982), reference frame selection (Meilinger et al., 2014, Werner and Schmidt, 1999) and switching costs between spatial units (Brockmole and Wang, 2002, Brockmole and Wang, 2003). Also interpretation of the current results clearly supports the theoretical distinction between VS and ES proposed by Montello (1993). Our results extend prior findings in that they demonstrate clear differences in the memory structure of different spaces on three different aspects by directly comparing VS and ES learning with the same material, thus, excluding additional differences.

A distinction between ES and VS based on opaque borders is found in neuroscience as well (for recent overviews on navigation see Spiers and Barry, 2015, Wolbers and Wiener, 2014). Visual borders were shown to influence the organization of spatial representations on the level of single neurons. Specialized cells fire whenever a rat is close to an enclosing wall (Solstad, Boccara, Kropff, Moser, & Moser, 2008) and opaque borders strongly influence the firing patterns of hippocampal place cells (O'Keefe & Burgess, 1996) as well as entorhinal grid cells (Stensola, Stensola, Moser, & Moser, 2015). A place cell fires at – and therefore identifies – a specific location within an environment (e.g., the south-west corner of a room). In a single room the same cell will show reactivation (in addition to base rate activity) if the same location is visited again. Importantly, across multiple interconnected spaces (i.e., within ES) often cells are not firing at a unique location only. Rather, the same cell may fire again (is reused) within different vista spaces (Grieves et al., 2016, Skaggs and McNaughton, 1998, Spiers et al., 2015). Transferred to the present experimentation, a single place cell might fire in multiple corridors, but it will not do so at corresponding locations within a single VS room. Not only place cells, but also grid cells are sensitive to compartmentalization along opaque VS borders (Carpenter, Manson, Jeffery, Burgess, & Barry, 2015). An entorhinal grid cell fires at repeated locations arranged along a regular grid covering the whole space. Interestingly, rats were found to use different grids for different corridors (ES), but a single grid pattern when walking similar trajectories within a single VS (Derdikman et al., 2009). These findings indicate that compartmentalization of ES into multiple VS along opaque borders is also reflected in the neuronal response.

When navigating towards a goal location, hippocampal place cells are activated consecutively along the route to that goal, even before physical movement (Pfeiffer & Foster, 2013). This successive activation has been proposed to correspond to mind (i.e., non-physical) travel – or mental walk – along a route (Byrne et al., 2007, Sanders et al., 2015). Path integration along mind travel may then be used to estimate a vector towards the goal. Indeed, such survey estimations were associated with hippocampal activity in humans before (Wolbers & Büchel, 2005). One specific prediction for path integration via mind travel is that longer distances towards a goal will result in more place cell activity and therefore larger overall hippocampal activation. Indeed changes in blood flow associated with higher summed activity at longer paths to a goal location was observed in humans as well, while watching a video of a travel through a familiar city part (Howard et al., 2014) and when sequentially presenting pictures of close-by and distant city locations (Morgan, Macevoy, Aguirre, & Epstein, 2011). It should, however, be noted that this increase of hippocampal activity can also be explained by an alternative process, namely, by mentally adding blocks of vista spaces to form a mental model of the non-visible surrounding (Meilinger, 2008). Importantly, this positive correlation of distance and hippocampal activity reversed when distances were introduced within a VS, i.e., a virtual room (Viard, Doeller, Hartley, Bird, & Burgess, 2011), or an endless plane (Sherrill et al., 2013). Thus, the human hippocampus presumably has a share in both spaces, but the processing involved differs qualitatively. In summary, the VS-ES distinction brought forward in the current study corresponds to some recent distinctions obtained from single cell activity in rodents as well as summed activity within humans. We do think that future experimentation along these lines will be fruitful.

Consistent with the literature our results show that memory for VS and ES differ due to visual borders. But how is that memory organized? A VS clearly seems to be treated as a unique unit. As in other studies where learning took place in a single room (e.g., Mou and McNamara, 2002, Shelton and McNamara, 2001) our results suggest a common reference frame for all locations in the VS conditions. Close-by and distant pointing targets were processed equally fast indicating similar access from within a common memory unit. While we do find reminiscence of the experienced order, this order effect still differs from ES learning and might be based on a memory system independent of the organization of spatial relations (Packard and McGaugh, 1996, Restle, 1957, van der Ham et al., 2010). Findings showing that place and grid cells in rats do not remap within a single constant VS (Derdikman et al., 2009, Skaggs and McNaughton, 1998) do further support the assumption of a single VS unit in memory. Similarly, this accounts for results demonstrating an advantage of mentally switching between object locations within a VS (Brockmole and Wang, 2002, Brockmole and Wang, 2003) and the preferred updating of object locations within a VS as compared to locations in neighbouring spaces

(Avraamides and Kelly, 2010, Kelly et al., 2007, Wang and Brockmole, 2003a, Wang and Brockmole, 2003b).

Contrary to VS, data from our and other studies suggest that ES memory is split into multiple units. Each unit, in our case, each individual corridor of the ES is assumed to operate as a VS. Thus, conclusions about underlying learning mechanisms drawn from the VS room should likewise be effective in a single corridor. In line with this, reference frame orientation followed the immediate visible input, both in a VS room as well as in an individual ES corridor. The observed distance effect in pointing latency suggests that memory access is fastest within the pre-activated memory unit, i.e., within the corridor one is currently located in. Beyond the visible unit the distance effect indicates a process of successive activation corridor per corridor, unit by unit, not a one-time recall of a single unit. Similar to switching costs, that are interpreted as an effortful retrieval of a new memory unit (Brockmole and Wang, 2002, Brockmole and Wang, 2003), latency increase with traveled distance can be interpreted as a successive activation (and integration) of the individually represented VS. This might be reflected in higher hippocampal activity with increasing path distance to the target location (Howard et al., 2014, Morgan et al., 2011). Our order effect in ES, which is exceeding the effect found under conditions of movement and successive presentation in VS, is also consistent with a structure of multiple connected subspaces which are accessed in the order of connection. Such a structure has already been proposed in the literature (Chrastil and Warren, 2014, Mallot and Basten, 2009, Meilinger, 2008, Trullier et al., 1997). Updating of object locations across subspaces might not naturally and easily emerge, as would be expected from single VS spatial unit. This is exactly what prior results showed (Avraamides and Kelly, 2010, Kelly et al., 2007, Wang and Brockmole, 2003a, Wang and Brockmole, 2003b). In sum, results from the present study as well as from the literature are consistent with the idea that ES are represented within multiple spatial units.

The representation of subspaces may rely on distinct reference frames, and the orientation of each reference frame might depend on the spatial cues available in each vista unit (Meilinger et al., 2014, Werner and Schmidt, 1999). These units might be further embedded within a hierarchical structure (Mallot and Basten, 2009, McNamara et al., 2008) with a common top-level reference frame encompassing multiple subunits. Such a top level reference frame might play a stronger role for individuals with high spatial abilities (e.g., Meilinger et al., 2014) or when familiarity with an environment increases. Importantly, extending the assumption of a single, common reference unit (e.g., Gallistel, 1990, O'Keefe and Nadel, 1978, Sholl, 2001) to ES without postulating a hierarchy with sublevels needs additional specification of the processes that generated the observed ES-specific distance and order effects and an explanation why these processes were not evoked in a VS.

The main conclusion from the present study is, that memory for VS and ES is structurally different – even if the same spatial information was learned. Accessing memory for ES was constrained by the distance and order in which objects were learned. We demonstrated that these effects cannot be fully explained by movement through the environment and successive object presentation, rather spatial separation is needed for that. The visible geometry of corridor and room determined the reference frame orientation in memory, and also likely the units ES memory was subdivided into. These results ultimately emphasize that transferring conclusions of findings obtained in VS studies to the more complex learning of ES (and vice versa) should be made cautiously.

Acknowledgements

This work was supported by the DFG grant ME 3476/2-2. We would like to thank Joachim Tesch and Betty Mohler for help with the virtual environment setup, Nadine Simon for help in running the experiments, and Michael Geuss for proof-reading.

Supplementary material

All data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.06.003.

References

Avraamides, M. N., & Kelly, J. W. (2010). Multiple systems of spatial memory: Evidence from described scenes. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(3), 635–645. http://dx.doi.org/10.1037/a0017040.

Brockmole, J. R., & Wang, R. F. (2002). Switching between environmental representations in memory. Cognition, 83(3), 295–316. http://dx.doi.org/10.1016/S0010-0277(02)00012-4.

Brockmole, J. R., & Wang, R. F. (2003). Changing perspective within and across environments. Brain & Development, 25(4), 291–293. http://dx.doi.org/10.1016/ S0.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. Psychological Review, 114(2), 340–375. http://dx.doi.org/10.1037/0033-295X.114.2.340.

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. http://dx.doi.org/10.1016/j.cub.2015.02.037.

Chrastil, E. R., & Warren, W. H. (2013). Active and passive spatial learning in human navigation: Acquisition of survey knowledge. Journal ofExperimental Psychology: Learning, Memory, and Cognition, 39(5), 1520–1537. http://dx.doi.org/10.1037/ a0032382.

Chrastil, E. R., & Warren, W. H. (2014). From cognitive maps to cognitive graphs. PLoS ONE, 9(11), e112544. http://dx.doi.org/10.1371/journal.pone.0112544.

Cohen, R., Baldwin, L. M., & Sherman, R. C. (1978). Cognitive maps of a naturalistic setting. Child Development, 49(4), 1216–1218. http://dx.doi.org/10.2307/ 1128763.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. Journal of Environmental Psychology, 24(3), 329–340. http://dx.doi.org/10.1016/ j.jenvp.2004.08.006.

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), 1325–1332. http://dx.doi.org/ 10.1038/nn.2396.

Fujita, N., Klatzky, R. L., Loomis, J. M., & Golledge, R. G. (1993). The encoding-error model of pathway completion without vision. Geographical Analysis, 25(4), 295–314. http://dx.doi.org/10.1111/j.1538-4632.1993.tb00300.x.

Gallistel, C. R. (1990). The organization of learning. Cambridge, MA: MIT Press. Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. Visual Cognition, 14(4–8), 704–715. http://dx.doi. org/10.1080/13506280500193719.

Greenauer, N., & Waller, D. (2010). Micro- and macroreference frames: Specifying the relations between spatial categories in memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(4), 938–957. http://dx.doi.org/ 10.1037/a0019647.

Grieves, R. M., Jenkins, B. W., Harland, B. C., Wood, E. R., & Dudchenko, P. A. (2016). Place field repetition and spatial learning in a multicompartment environment. Hippocampus, 26(1), 118–134. http://dx.doi.org/10.1002/hipo.22496.

Hardiess, G., Gillner, S., & Mallot, H. A. (2008). Head and eye movements and the role of memory limitations in a visual search paradigm. Journal of Vision, 8(1), 7.1–13. http://dx.doi.org/10.1167/8.1.7.

Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., ... Spiers, H. J. (2014). The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. Current Biology, 24(12), 1331–1340. http:// dx.doi.org/10.1016/j.cub.2014.05.001.

Janzen, G. (2006). Memory for object location and route direction in virtual large- scale space. The Quarterly Journal of Experimental Psychology, 59(3), 493–508. http://dx.doi.org/10.1080/02724980443000746.

Kelly, J. W., Avraamides, M. N., & Loomis, J. M. (2007). Sensorimotor alignment effects in the learning environment and in novel environments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(6), 1092–1107. http://dx.doi.org/10.1037/0278-7393.33.6.1092.

Kelly, J. W., & McNamara, T. P. (2008). Spatial memories of virtual environments: How egocentric experience, intrinsic structure, and extrinsic structure interact. Psychonomic Bulletin & Review, 15(2), 322–327. http://dx.doi.org/10.3758/ PBR.15.2.322.

Klatzky, R. L., Loomis, J. M., Golledge, R. G., Cicinelli, J. G., Doherty, S., & Pellegrino, J. W. (1990). Acquisition of route and survey knowledge in the absence of vision. Journal of Motor Behavior, 22(1), 19–43. http://dx.doi.org/10.1080/ 00222895.1990.10735500.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. Journal of Experimental Psychology: Human Perception and Performance, 4(1), 47–60. http://dx.doi.org/ 10.1037/0096-1523.4.1.47.

Kosslyn, S. M., Pick, H. L., & Fariello, G. R. (1974). Cognitive maps in children and men. Child Development, 45(3), 707–716. http://dx.doi.org/10.1111/1467-8624. ep12147012.

Levin, J. R., & Neumann, E. (1999). Testing for predicted patterns: When interest in the whole is greater than in some of its parts. Psychological Methods, 4(1), 44–57. http://dx.doi.org/10.1037//1082-989X.4.1.44.

Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. a. (1993). Nonvisual navigation by blind and sighted: Assessment of path integration ability. Journal of Experimental Psychology: General, 122(1), 73–91. http://dx.doi.org/10.1037/0096-3445.122.1.73.

Mallot, H., & Basten, K. (2009). Embodied spatial cognition: Biological and artificial systems. Image and Vision Computing, 27(11), 1658–1670. http://dx.doi.org/ 10.1016/j.imavis.2008.09.001.

McNamara, T. P. (1986). Mental representations of spatial relations. Cognitive Psychology, 18(1), 87–121. http://dx.doi.org/10.1016/0010-0285(86)90016-2.

McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In Spatial cognition III (pp. 174–191). Berlin, Heidelberg: Springer Berlin Heidelberg. http://dx.doi.org/10.1007/3-540-45004-1_11.

McNamara, T. P., Sluzenski, J., & Rump, M. (2008). Human spatial memory and navigation. In H. L. Roediger, III (Ed.). Cognitive psychology of memory. Learning and memory: A comprehensive reference (Vol. 2, pp. 157–178). Oxford: Elsevier.

Meilinger, T. (2008). The network of reference frames theory: A synthesis of graphs and cognitive maps. Spatial cognition VI. Learning, reasoning, and talking about space (Vol. 5248, pp. 344–360). Berlin, Heidelberg: Springer. http://dx.doi.org/ 10.1007/978-3-540-87601-4_25.

Meilinger, T., & Bülthoff, H. H. (2013). Verbal shadowing and visual interference in spatial memory. PLoS ONE, 8(9), e74177. http://dx.doi.org/10.1371/journal. pone.0074177.

Meilinger, T., Frankenstein, J., Watanabe, K., Bülthoff, H. H., & Hölscher, C. (2015). Reference frames in learning from maps and navigation. Psychological Research, 79(6), 1000–1008. http://dx.doi.org/10.1007/s00426-014-0629-6.

Meilinger, T., Knauff, M., & Bülthoff, H. (2008). Working memory in wayfinding—A dual task experiment in a virtual city. Cognitive Science: A Multidisciplinary Journal, 32(4), 755–770. http://dx.doi.org/10.1080/03640210802067004.

Meilinger, T., Riecke, B. E., & Bülthoff, H. H. (2014). Local and global reference frames for environmental spaces. Quarterly Journal of Experimental Psychology (2006), 67(3), 542–569. http://dx.doi.org/10.1080/17470218.2013.821145.

Moar, I., & Carleton, L. R. (1982). Memory for routes. The Quarterly Journal of Experimental Psychology Section A, 34(3), 381–394. http://dx.doi.org/10.1080/ 14640748208400850.

Montello, D. R. (1993). Scale and multiple psychologies of space. Spatial Information Theory, 312–321. http://dx.doi.org/10.1007/3-540-57207-4_21.

Morgan, L. K., Macevoy, S. P., Aguirre, G. K., & Epstein, R. A. (2011). Distances between real-world locations are represented in the human hippocampus. The Journal of Neuroscience, 31(4), 1238–1245. http://dx.doi.org/10.1523/ JNEUROSCI.4667-10.2011.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(1), 162–170. http://dx.doi.org/10.1037//0278-7393.28.1.162.

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(1), 142–157. http://dx.doi.org/10.1037/ 0278-7393.30.1.142.

Newcombe, N., & Liben, L. S. (1982). Barrier effects in the cognitive maps of children and adults. Journal of Experimental Child Psychology, 34(1), 46–58. http://dx.doi. org/10.1016/0022-0965(82)90030-3.

O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. Nature, 382(6581), 425–428.

O'Keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. Oxford: Oxford University Press.

Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiology of Learning and Memory, 65(1), 65–72. http://dx.doi.org/ 10.1006/nlme.1996.0007.

Pantelides, S. N., Kelly, J. W., & Avraamides, M. N. (2016). Integration of spatial information across vision and language. Journal of Cognitive Psychology, 28(2), 171–185. http://dx.doi.org/10.1080/20445911.2015.1102144.

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. Nature, 497(7447), 1–8. http://dx.doi.org/ 10.1038/nature12112.

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. Journal ofExperimental Psychology: General, 134(1), 73–92. http://dx.doi.org/10.1037/0096-3445.134.1.73.

Restle, F. (1957). Discrimination of cues in mazes: A resolution of the “place-vs.- response” question. Psychological Review, 64(4), 217–228. http://dx.doi.org/ 10.1037/h0040678.

Sanders, H., Rennó-Costa, C., Idiart, M., & Lisman, J. E. (2015). Grid cells and place cells: An integrated view of their navigational/memory function. Trends in Neurosciences, 38(12), 763–775. http://dx.doi.org/10.1016/j.tins.2015.10.004.

Schweizer, K., Herrmann, T., Janzen, G., & Katz, S. (1998). The route direction effect and its constraints. Spatial cognition (Vol. 15, pp. 19–38). . http://dx.doi.org/ 10.1007/3-540-69342-4_2.

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. Cognitive Psychology, 43(4), 274–310. http://dx.doi.org/10.1006/ cogp.2001.0758.

Sherrill, K. R., Erdem, U. M., Ross, R. S., Brown, T. I., Hasselmo, M. E., & Stern, C. E. (2013). Hippocampus and retrosplenial cortex combine path integration signals for successful navigation. Journal of Neuroscience, 33(49), 19304–19313. http:// dx.doi.org/10.1523/JNEUROSCI.1825-13.2013.

Sholl, M. J. (2001). The role of a self-reference system in spatial navigation. In D. R. Montello (Ed.), Cosit 2001 (pp. 217–232). Berlin, Heidelberg: Springer. http://dx. doi.org/10.1007/3-540-45424-1_15.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. Advances in Child Development and Behavior, 10, 9–55. http://dx.doi.org/10.1016/S0065-2407(08)60007-5.

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.

Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., & Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. Science, 322 (5909), 1865–1868. http://dx.doi.org/10.1126/science.1166466.

Spiers, H. J., & Barry, C. (2015). Neural systems supporting navigation. Current Opinion in Behavioral Sciences, 1, 47–55. http://dx.doi.org/10.1016/ j.cobeha.2014.08.005.

Spiers, H. J., Hayman, R. M. a., Jovalekic, A., Marozzi, E., & Jeffery, K. J. (2015). Place field repetition and purely local remapping in a multicompartment environment. Cerebral Cortex, 25(1), 10–25. http://dx.doi.org/10.1093/cercor/bht198.

Stensola, T., Stensola, H., Moser, M.-B., & Moser, E. I. (2015). Shearing-induced asymmetry in entorhinal grid cells. Nature, 518(7538), 207–212. http://dx.doi. org/10.1038/nature14151.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 57(4), 421–457.

Street, W. N., & Wang, R. F. (2014). Differentiating spatial memory from spatial transformations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(2), 602–608. http://dx.doi.org/10.1037/a0035279.

Strickrodt, M., O'Malley, M., & Wiener, J. M. (2015). This place looks familiar—How navigators distinguish places with ambiguous landmark objects when learning novel routes. Frontiers in Psychology, 6(1936), 1–12. http://dx.doi.org/10.3389/ fpsyg.2015.01936.

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. Cognitive Psychology, 14(4), 560–589. http://dx.doi.org/10.1016/0010-0285(82)90019-6.

Tlauka, M., Carter, P., Mahlberg, T., & Wilson, P. N. (2011). The first-perspective alignment effect: The role of environmental complexity and familiarity with surroundings. The Quarterly Journal of Experimental Psychology, 64(11), 2236–2250. http://dx.doi.org/10.1080/17470218.2011.586710.

Trullier, O., Wiener, S. I., Berthoz, A., & Meyer, J. A. (1997). Biologically based artificial navigation systems: Review and prospects. Progress in Neurobiology, 51 (5), 483–544. http://dx.doi.org/10.1016/S0301-0082(96)00060-3.

Valiquette, C., & McNamara, T. P. (2007). Different mental representations for place recognition and goal localization. Psychonomic Bulletin & Review, 14(4), 676–680. http://dx.doi.org/10.3758/BF03196820.

van der Ham, I. J. M., van Zandvoort, M. J. E., Meilinger, T., Bosch, S. E., Kant, N., & Postma, A. (2010). Spatial and temporal aspects of navigation in two neurological patients. NeuroReport, 21(10), 685–689. http://dx.doi.org/ 10.1097/WNR.0b013e32833aea78.

Viard, A., Doeller, C. F., Hartley, T., Bird, C. M., & Burgess, N. (2011). Anterior hippocampus and goal-directed spatial decision making. Journal of-Neuroscience, 31(12), 4613–4621. http://dx.doi.org/10.1523/JNEUROSCI.4640-10.2011.

Waller, D., Loomis, J. M., & Haun, D. B. M. (2004). Body-based senses enhance knowledge of directions in large-scale environments. Psychonomic Bulletin & Review, 11(1), 157–163. http://dx.doi.org/10.3758/BF03206476.

Wan, X., Wang, R. F., & Crowell, J. a. (2013). Effects of basic path properties on human path integration. Spatial Cognition & Computation, 13(1), 79–101. http:// dx.doi.org/10.1080/13875868.2012.678521.

Wang, R. F., & Brockmole, J. R. (2003a). Human navigation in nested environments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(3), 398–404. http://dx.doi.org/10.1037/0278-7393.29.3.398.

Wang, R. F., & Brockmole, J. R. (2003b). Simultaneous spatial updating in nested environments. Psychonomic Bulletin &Review, 10(4), 981–986. http://dx.doi.org/ 10.3758/BF03196562.

Wen, W., Ishikawa, T., & Sato, T. (2011). Working memory in spatial knowledge acquisition: Differences in encoding processes and sense of direction. Applied Cognitive Psychology, 25(4), 654–662. http://dx.doi.org/10.1002/acp.1737.

Werner, S., & Schmidt, K. (1999). Environmental reference systems for large-scale spaces. Spatial Cognition and Computation, 1, 447–473. http://dx.doi.org/ 10.1023/A:1010095831166.

Wiener, J. M., Kmecova, H., & de Condappa, O. (2012). Route repetition and route retracing: Effects of cognitive aging. Frontiers in Aging Neuroscience, 4, 1–7. http://dx.doi.org/10.3389/fnagi.2012.00007.

Wiener, J. M., & Mallot, H. A. (2006). Path complexity does not impair visual path integration. Spatial Cognition & Computation, 6(4), 333–346. http://dx.doi.org/ 10.1207/s15427633scc0604_3.

Wilson, P. N., Wilson, D. a., Griffiths, L., & Fox, S. (2007). First-perspective spatial alignment effects from real-world exploration. Memory & Cognition, 35(6), 1432–1444. http://dx.doi.org/10.3758/BF03193613.

Wolbers, T., & Büchel, C. (2005). Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. Journal of Neuroscience, 25(13), 3333–3340. http://dx.doi.org/10.1523/JNEUROSCI.4705- 04.2005.

Wolbers, T., & Wiener, J. M. (2014). Challenges for identifying the neural mechanisms that support spatial navigation: The impact of spatial scale. Frontiers in Human Neuroscience, 8, 1–12. http://dx.doi.org/10.3389/ fnhum.2014.00571.

Yamamoto, N., & Shelton, A. (2009). Sequential versus simultaneous viewing of an environment: Effects of focal attention to individual object locations on visual spatial learning. Visual Cognition, 17(4), 457–483. http://dx.doi.org/10.1080/ 13506280701653644.

Yamamoto, N., & Shelton, A. L. (2005). Visual and proprioceptive representations in spatial memory. Memory & Cognition, 33(1), 140–150. http://dx.doi.org/ 10.3758/BF03195304.

Yamamoto, N., & Shelton, A. L. (2007). Path information effects in visual and proprioceptive spatial learning. Acta Psychologica, 125(3), 346–360. http://dx. doi.org/10.1016/j.actpsy.2006.09.001.

7.2 Study 2: Routes embedded in survey knowledge

Study reference

Title: Spatial Survey Estimation Is Incremental and Relies on Directed Memory Structures

Authors: Tobias Meilinger, Marianne Strickrodt, Heinrich H. Bülthoff

Status: Accepted, Published

How to cite: Meilinger, T., Strickrodt, M., & Bülthoff, H.H. (2018). Spatial Survey Estimation Is Incremental and Relies on Directed Memory Structures. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI. 11th International Conference, Spatial Cognition 2018. Lecture Notes in Artificial Intelligence* (Vol. 11034). Springer.

Abstract

This study examined how navigators of large-scale environmental spaces come up with survey estimates of distant targets. Participants learned a route through a virtual city by walking it multiple times in one direction on an omnidirectional treadmill. After learning, they were teleported to intersections along the route and pointed to multiple other locations. Locations were always queried in chunks of related trials relative to a participant's current position, either to all locations route forwards or all locations route backwards. For their first pointing, participants took twice as long as for the later pointings and latency correlated with the number of intersections to the target, which was not the case for later pointings. These findings are inconsistent with reading out coordinates from a cognitive map but fit well with constructive theories which suggest that participants integrated locations between their current location and the target along the learned path. Later pointings to adjacent intersections within a chunk of trials continued this process using the previous estimation. Additionally, in first pointings participants' estimates were quicker and more accurate when targets were located route forwards than route backwards. This route direction effect shows that the long-term memory employed in generating survey estimates must be directed – either in form of a directed graph or a combination of a directed route layer and an undirected survey layer.

Keywords: spatial memory; survey knowledge; environmental space; cognitive map; mental walk; mental model; virtual environment

Introduction

After walking through cities and buildings humans can grasp metric relationships such as distances and directions between remote landmarks. In order to do so they must integrate spatial information obtained across multiple views and places along their navigation trajectory. How do humans store the experienced information and how do they infer survey relations from them when asked to do so?

To solve a survey task, such as pointing to a distant landmark, at least one's current location and the target location must be brought into direct reference. Some theories assume that navigators form a global, world-centered reference frame within which all relevant locations are represented [1–4]. In the following a global world-centered reference frame will be called a cognitive map. Survey relations can be obtained from a cognitive map by *reading out the coordinates* of the relevant locations (e.g., the current location and the target location) and compute the difference vector between the coordinates to get the relative direction or the distance between the locations, etc. An alternative approach is taken by theories suggesting that a navigable space is not represented within a cognitive map, but by multiple local memory units which are connected in a graph structure [5–7]. For such graph structures Meilinger [7] suggested that for making survey estimates the integration of one's current position and the target within a single reference frame happens on the fly during retrieval by constructing a *mental model* of the non-visible environment (a related vector-addition model was presented for updating by Fujita et al. [8]). For example, navigators could imagine what the environment would look like if the surrounding walls were transparent. First, they imagine the adjacent street from their current position, then they add the street branching off from it, etc. In this way all locations from the current location along a route leading towards the target location are imagined step by step within the current egocentric reference frame, building a mental model of the environment. No one mentally walks through this constructed environment and the underlying memory structure is no cognitive map, but a graph consisting of local memory units of places interconnected by links.

Increasing evidence for the presence of local memory units can be found in the literature. The use of multiple, locally confined reference frames (one for each corridor) for pointing to distant targets was shown in multi-corridor environments [9–11]. Also, knowledge of spatial relations of targets within a single room seems to be partly dissociated from the knowledge about the location of the room itself [12]. Those studies clearly support graph theories [5–7]. In several studies, longer reaction times were shown for recalling a target location the more local units (e.g., individual corridors) were experienced along the path during learning between one's current location and the target [11,13,14]. In our study we will refer to this effect as the “distance effect”, which should therefore not be understood in its Euclidean sense (i.e., straight line distance), but instead refers to the number of locations visited along a route. Such distance effects can be well explained by a mental model built from a graph-like memory structure [7]. Here, a time consuming, incremental process of activating spatial information along the learned route is underlying the estimation of the relative direction of a target. In contrast, there are other studies supporting the idea of global, cognitive maps, which could be used for a simple read out of coordinates. For example, some studies indicate that participants form reference frames (or reference directions) that are covering multiple local subspaces, such as corridors or streets [9, 11, 15]. They suggest that all spatial information gathered across multiple subspaces have been stored (also) relative to a single reference system in long-term memory. Furthermore, several models allow [6] or propose [16] the combination of local (often route related) and global (typically survey related) memory structures.

Many empirical findings suggest that human spatial memory is directed, or in other words, asymmetric. For example, people occasionally select different routes when either going from A to B compared to going from B to A [17]. Also, the error patterns that are observed when participants estimate the relative direction along a route from location A to location B do not coincide with error patterns when pointing from B to A ([14] same volume). This indicates that no coherent map was underlying survey performance. Two propositions seem eligible to account for such results. Either, one could argue that two (rather than one) coherent cognitive maps have been built, one for the forward one for the backward direction of the route, which do not need to coincide. No additional information about directed connections between locations need to be stored. Depending on the direction queried (i.e., either from A to B or from B to A), either of the two maps is selected, leading to the observed asymmetries in pointing directions. Alternatively, no global embedding took place, but a graph structure with local memory units that are connected by directed links was stored, for example, a link emphasizing the direction from A to B, but not the other way around [6, 7]. The directed links might render different paths to be preferred for forward and backward route planning and may lead to asymmetric pointing errors as link usage along the link orientation is easier than in opposite direction. For the latter the link must be inverted, which is computationally costly. Support for an embedded directedness in spatial memory comes from studies utilizing primed recognition of landmarks. Re-cognizing landmarks previously experienced along a route is faster when they are preceded by another landmark in the same order as during learning, compared to being preceded by a landmark that was succeeding the target during learning [18, 20]. This route direction effect is explained by a directed encoding of connected places in the experienced direction. However, it is unclear whether findings obtained from this simple recognition task generalize to survey tasks as well. Therefore, we aimed to investigate whether directedness is a determinative part of a large-scale space representation utilized in a survey task. To exclude the possibility that navigators formed separate memory structures for route forwards and backward learning (as done by [17] same volume) we had participants learn a route only in one direction.

An interesting aspect not yet addressed in the literature is the question of how transient constructed survey estimations in working memory are and whether subsequent survey estimations can be based on them. Imagine learning landmarks A, B, C, and *d* along a route and being queried the bearing of *d* while standing at A in a first trial. Following a construction model, location B and C would be successively activated on the way of mentally walking to or constructing the relative location of D. Now, having pointed to *d* you are subsequently asked to point to C, the direct neighbor of D. Either this can be done by again constructing a new model from A via B to C. Alternatively, subsequent pointing to C could also be based on the previous estimation of *d* and calculating backwards from there to derive the location of C. In short, one could use information from the old model to compute subsequent steps from there rather than built a new model from scratch. In that case later pointings should be much quicker than first pointings if neighboring targets are queried, and their latency should not depend on the distance between pointer and target. In our study we set out to examine whether the recall of survey relations is based on all-at-once or incremental processes and whether prior recall of related locations can serve as a base for succeeding targets.

Experiment and Predictions

We had participants learn a virtual route containing a set of to-be-learned locations multiple times from start to end. Subsequently, we administered a survey task where participants were teleported to different intersections along the route and needed to face straight line direction towards several of the remaining intersections. Hereby, we manipulated multiple factors. We always queried chunks of related locations. Being teleported to an intersection, participants always had to successively recall a sequence of neighboring intersections. This was administered to examine whether later pointing was influenced by prior pointing estimates. Furthermore, targets were always selected relative to participants' current location on the route following two rules: Firstly, we varied whether the targets were lying towards the end of the route (i.e., forward, in route direction) or towards the start of the route (i.e., backwards, against route direction). Secondly, in order to balance the number of intersections between current location and target (i.e., route distance) for first and later pointings within a chunk of related trials, participants were pointing to locations in a target sequence either away from their current location (i.e., first a minimum distance to the adjacent intersection, then the second-next intersection, etc.) or participants pointed in a sequence towards their current location (i.e., first a maximum distance to the start or the end of the route, then to the second/second last intersection, etc., until ending up pointing to the neighboring intersection).

Depending on the underlying memory structure and retrieval process different predictions can be made. An all-at-once read-out process from a cognitive map would neither predict an effect of distance to the target nor an effect of route direction on the performance in the direction estimation task. In contrast, a graph representation accessed via the construction of a mental model assumes a time-consuming incremental retrieval of survey knowledge along the successively visited places towards the target, thus, taking the longer the further the target is away from the navigator along the route (distance effect). Additionally, if the graph representation consists of directed links between adjacent places faster recall of targets located towards the end of the route relative to one's current position should be shown (in learned route direction) compared to estimating direction to targets located towards the start of the route (against route direction). Regarding the interdependence between trials, later pointings within a chunk of trials may re-iterate the whole process and yield identical results as initial pointings. Alternatively, participants may build upon earlier pointing estimates and only add the difference from the previous target to the adjacent intersection. In this case later pointings should be quicker than earlier pointings, show no distance effect and route directions effects might cancel out each other as later pointings depending on the target sequence (towards or away) follow equally often a route upwards and downwards direction.

Methods

Participants

24 participants took part in the experiment. One participant's performance did not significantly differ from chance and was not included, leading to 23 participants (11 females and 12 males) aged between 21 and 64 (*M* = 29.6 years, *SD* = 9.3 years) used in the analysis. All participants were recruited via a subject-database, gave written informed consent, and were paid for their participation. The procedure was approved by the ethical committee of the University Clinics Tübingen.

Fig. 1. The virtual city as seen from navigation perspective (left side) and from bird's eye view with the route marked in red (right side). During learning the start, the end and each of the six intersections in-between were marked with white crosses on the floor (marked by red dots in this figure). They served as locations to be teleported to and as targets during the test phase.

Material

The Virtual City

In the learning phase, participants had to learn a route through a virtual city. Figure 1 shows a snapshot of the city as seen during walking, as well as a bird's eye view of the route. The route consisted of a start, six intersections and an end, resulting in eight locations that served as targets during testing. During learning, all eight locations were marked with a white X on the floor. The type of houses changed along the route, as did street width and the heights of houses. In addition, individual houses ensured sufficient landmark information to identify each location. The eight locations were not labelled by names.

The Setup

Participants walked on a 4x4 meters omnidirectional treadmill (Fig. 2 left side). It allowed them to walk for infinite distances in any direction by moving them back to the center of the treadmill. This unique interface allows for realistic proprioceptive and vestibular feedback as well as efference copies while walking in virtual environments. Participants wore a climbing harness for the unlikely event of falling and hurting themselves on the moving platform. To obtain participants' location on the treadmill we tracked their head position with 16 high-speed motion capture cameras at 120 Hz (Vicon® MX 13). This data was used both to control the treadmill and to update the visualization of the virtual environment. The visual surrounding at a location was rendered in real time (60Hz) using a NVIDIA Quadro FX 4600 graphics card with 768 MB RAM in a standard PC. Cables connected the PC to the display via the ceiling. Participants viewed the scene in stereo using a nVisor SX60 head-mounted display that provided a field of view of 44×35 degrees at a resolution of 1280×1024 pixels for each eye with 100% overlap. The setup thus also provided important visual depth cues such as stereo images and motion parallax. During the test phase a circular handrail around them with 0.48 meters diameter prevented participants from leaving their location (Fig. 2 middle) and responses were given by rotating the head and pressing a button on a gamepad they were holding.

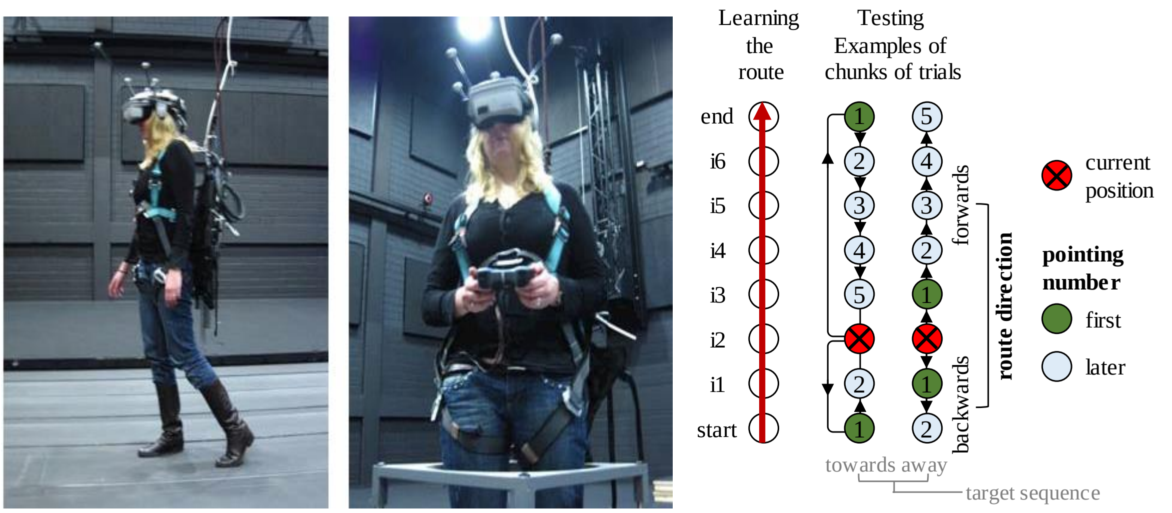


Fig. 2. Left: Participant walking on the omnidirectional treadmill during learning. Middle: Participant pointing to a target during testing by facing the target and pressing a button on a gamepad. Right: Order of learning and examples of chunks of testing phase. The current position could be at any of the eight locations, *distance* to the targets would vary accordingly. Factors *pointing number* and *route direction* are visualized and that we varied the target sequence.

Procedure

In the learning phase, participants walked the route at least six times from start to end. They were instructed to first learn the route, and secondly be able to self-localize when teleported to an X along the route after the learning phase. Participants were free to look around as long as they wanted, however they were not allowed to look or walk back to where they came from. In their first run, they walked up to an intersection, looked around, and the experimenter pointed out the street to take when the participant looked down the correct street by stating “the route is this direction” (the experimenter was in the same room and could talk with the participant). No verbal turning information (e.g., “left”, “straight on”, etc.) was given. When reaching the end and having looked around participants were teleported back to the start. From the second run onwards participants were asked to approach an intersection, look into the direction the route was going on and say “this way”. The experimenter gave feedback whether this was right or wrong, before participants proceeded. They were not allowed to leave the route. For each new run, the virtual environment was rotated 90° clockwise relative to the lab. Sound sources within the lab could thus not be used to derive global orientation. The learning phase ended when participants walked the route at least six times and at least two runs were error-free. This criterion ensured comparable levels of route knowledge for all participants. Participants briefly trained walking on the treadmill before starting the experiment.

In the following test phase, participants were teleported to the eight locations on the route (i.e., the start, the end or i1-i6). The mark (i.e., X) for all locations was removed. For self-localization, participants could look and rotate around, but not walk around. As soon as they subjectively knew their location and orientation, they were asked to press a button on a gamepad. Then they pointed to a chunk of multiple targets. Pointing was done by turning on the spot until a vertical black line in the middle of the display matched the direction in which the participant thought the target was located. Thus, they would look directly at the target location as if the surrounding houses were transparent. When participants thought they faced the target, they pressed a button to confirm the direction and then pointed to the next target. No feedback was provided. After they had pointed to all targets within a chunk, participants pressed a second button on the gamepad and were teleported to a new position.

Figure 2 right, visualizes examples for four chunks of trials participants had to solve. The initial trial within each chunk was labelled “first” trial (dark green in figure), the remaining as “later” trials (light blue in figure), yielding the factor *pointing number*, which was introduced to examine potential dependencies between subsequent survey estimates. Four conditions determined the targets and the order in which participants were asked to point towards them within a chunk of trials. For each chunk they were instructed to point either (1) first to the start and then to all locations between start and their current location in the order of walking (i.e., start, i1, i2, etc.) (lower left example in figure), or (2) they should point to the same locations, but in reverse order (i.e., first the intersection before the current location, then the intersection before that, etc. until finally pointing to the start) (lower right example in figure). (3) They should point to the next intersection along the learned route direction after their current location, then the second next, etc. until pointing to the end (upper right example in figure). Or they should (4) point first to the end, then i6, i5, etc. until pointing to the intersection after their current location (upper left example in figure). Consequently, we varied the *route direction* (backwards to start vs. forwards to end) and target sequence (away vs. towards the current location) within a chunk of trials. *Route direction* served as a factor for analyzing potential directedness in survey estimates. Target sequence was introduced to balance average distance from the current location for first and later targets. Depending on one's current location along the route the maximum number of intersections one had to point to within a chunk of trials varied. For example, as visualized in Figure 2 right, standing at i2 facing backwards to the start involves two targets to point to with decreasing or increasing route leg distance across a chunk depending on being queried in towards or away target sequence, while facing forward to the end involves five targets/intersections to point to. Therefore, *distance* in terms of route legs varied across the experimental trials. Please note that the adjacent, neighboring intersections were always visible during pointing.

From the eight locations on the route (including start and end) participants pointed to every other location twice (away and towards their current location). The 28 pointing chunks were presented in random order for each participant (pointing route forwards from seven locations, backwards from seven locations, both in two target sequences). This whole procedure was repeated resulting in 56 pointing chunks. After finishing a chunk participants received feedback on how many targets they missed or how many redundant targets they added. No feedback about pointing accuracy was provided. Chunks with too few or too many responses were not analyzed as target locations could not be assigned. We recorded self-localization time (not reported), pointing latency and pointing direction for each trial. For the analysis we used latency and computed the absolute pointing error (i.e., the deviation between correct and estimated pointing direction irrespective of the direction of the error). Values deviating more than three standard deviations from the overall mean were not analyzed. Individual pointing accuracy all differed significantly from a random pointing behavior (i.e., 90°), indicating that all participants acquired some survey knowledge.

Results

To ascertain potential directedness in survey estimates as well as dependencies between subsequent pointings we first conducted a 2 x 2 ANOVA with the factors *pointing number* (first vs. later pointings) and *route direction* (pointing route forwards towards end vs. route backwards towards start). Table 1 summarizes the results for this analysis, Figure 3 visualizes the performance patterns. Both main factors show a significant effect on latency, *pointing number* also on error. Additionally, both for error and latency, *pointing number* interacted significantly with *route direction*.[13] For first pointings participants pointed quicker (t=4.01, p<.001) and more accurately (t=2.11, p=.042) when the target was located route forwards towards the end than when located route backwards towards the start. This indicates a route direction effect in survey estimates predicted by directed graph models, but is not expected when reading out coordinates from a cognitive map. Interestingly, no such differences occurred in later pointings (t's< 1.2, p's >.23). Participants pointed slower in their first pointing than for later pointings, but also conducted less errors. The effect on latency is consistent with incremental graph theories when assuming that subsequent estimates are based on previous estimates to their direct neighbors.

Table 1. Results of the ANOVA for latency and error. Degrees of freedom are $F(1, 22)$ for each factor(-combination). Significant effects are marked in bold.

	Latency			Error		
	F	p	η_p^2	F	p	η_p^2
Pointing number	**51.23**	**< .001**	**.70**	**34.86**	**< .001**	**.61**
Route direction	**10.04**	**.004**	**.31**	0.27	.608	.01
Number x direction	**6.50**	**.018**	**.23**	**11.79**	**.002**	**.35**

[13] When including target sequence (albeit not decisive on the introduced models) into the analysis all reported effects remained significant. There was no significant three-way-interaction which could have changed one of the reported effects, and no interaction with route direction. The analysis showed an effect of target sequence and its interaction with pointing number. Here participants were much quicker and accurate when their first pointing was away from their current location towards the visible neighbor intersection.

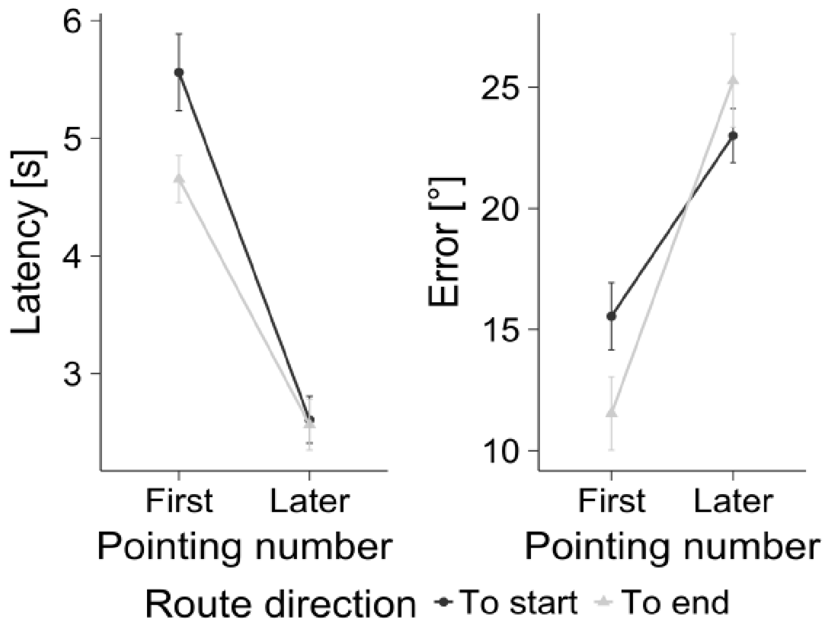


Fig. 3. Pointing performance in the form of latency (left) and absolute error (right) as a function of route direction and pointing number. Means and within-participants standard errors as estimated from the marginal means are shown.

For the further investigation of a potential incremental process of recalling survey knowledge we additionally considered a correlation analysis, namely, we examined whether latency and error for first and later pointings were associated with the route leg *distance* to the target. Indeed, first pointings showed a *distance* effect on latency. The further away along the route path the target was the longer participants required for pointing as indicated in a positive correlation between distance to the target and latency, with an average correlation of r=0.39, SD=0.29, significantly larger than zero, $t(22)$=6.55, p<.001. We observed no such correlation for later pointings, r=0.03, SD=0.12, difference from zero $t(22)$=1.03, p=.31. Errors correlated with distance both for first pointings, r=0.70, SD=.14, $t(22)$=23.8, p<.001, as well as for later pointings, r=0.45, SD=0.19, $t(22)$=11.5, p<.001.

Discussion

In our study we aimed to clarify whether survey estimates within navigable space are based on the incremental process of recalling target locations from a graph representation including the successive place-to-place activation of spatial information along the learned path, or whether they are based on an immediate read out of coordinates from an integrated cognitive map. More precisely, in case of reliance on a graph representation we examined whether survey estimates are based on directional encoding in long-term memory and also whether subsequent survey estimates will depend on previous estimates, thus, continuing the incremental process of recalling place-to-place information.

The Route Direction Effect in Survey Knowledge

We found a route direction effect, namely, a difference in performance for first pointing trials depending on whether participants pointed to the start or the end of the route. Participants pointed quicker and more accurate to targets located route forward towards the end of the route compared to pointing route backwards towards the start during the initial trial within a chunk. Such results support graph theories that assume route forwards encoding, as this directed encoding should speed-up integration towards the end, but slow down integration towards the start. Results for the first pointing are in line with asymmetries observed in spatial memory before in landmark recognition [18, 20] and route choice [17] and extend them to survey tasks (see also [14] same volume). They suggest that participants' long-term memory consisted of a directed graph and survey estimates were directly constructed from that graph. Such a directed graph was proposed by Meilinger [7] and the undirected graph-model from Chrastil and Warren [5] is easily adjusted to it. The effect of route direction was not present in later pointings, indicating interdependence of successions of trials discussed further in the following section.

Incremental Integration and Interdependence of Survey Estimates

Contrasting incremental graph theories with simple read-out of coordinates from a cognitive map, only the former predict a route leg distance effect for latency (processing speed) in the first, but not in later pointings of a chunk of trials and interdependence between trials within a chunk (i.e., faster later pointings which profit from earlier pointings). Consistent with incremental theories participants pointed slower in their first pointing than for later pointings, but also conducted less errors. For the first pointing within a chunk participants had to integrate all intersections between the current location and the target. This time took the longer the more intersections were involved as indicated in the positive correlation of latency with route leg distance—an indicator for successive activation of local memory units along the previously learned path, rather than a read out from a cognitive map. Later pointings, successively following neighboring locations of the previous target, showed different patterns. Performance was much quicker on average and did not correlate with the target distance from the participant. This suggests that participants did not repeat the incremental process of integrating all intersections between their current location and the target again, but only added or subtracted the single segment between the old and the new target. Targets for first pointings were on average 2.5 intersections away (averaging towards and away target sequence, where initial pointings for away chunks have a route leg distance of one intersection and where initial pointings for towards chunks can vary between one and seven). In the case of interdependence between trials in a chunk later pointings are always just one intersection away from the previous estimate. Thus, the mean difficulty for estimating the direction to a new target with regards to a distance effect is lower for later pointings compared to first pointings and quicker reaction expected. Alternatively, no new estimate had to be conducted, but instead the target was already present in working memory as part of the constructed mental model and just had to be accessed from there. Neither effect on latency would be expected by read-out from a cognitive map.

For error both in first and later pointings route leg distance correlated with error. This could be due to errors encoded in long-term memory. Assuming a roughly constant random error during encoding, integration across larger distances will aggregate larger errors no matter which process is used. In fact, all models would assume such an effect. In case of integration into a cognitive map, this map would store all locations inside a single reference system, but in a distorted way. In addition to the overall distance effect, error was larger for later pointings. A simple all-at-once read-out from a cognitive map would not predict such a difference, but incremental models do so. In line with latency results, building upon first estimates, adds up the number of estimates across the chunk of trial. Higher error can be explained by assuming additional error for every mental processing step that is made.

Please note that longer latency for first pointings cannot be explained by additional processing time for self-localization as this happened before pointing. Another aspect is the required head turn. For the first pointing one can expect an average turn of 90° (from a random heading during self-localization to first target). For later pointings, participants only turned towards an adjacent intersection which required a clearly smaller average head turn. We reckon that head movement itself surely is a part of the overall performance but that the observed average latency difference of 2.6 seconds between first and later estimates encompasses other processes as well. Furthermore, head turning cannot explain the distance effect in our experiment (distant targets do not necessarily require larger head turns – see Fig. 1) and other experiments where no head turning was involved at all and distance effects were still observed [11, 13]. Participants took longer and were more accurate for first pointings, but quicker and more error prone for later pointings. As latency and error correlated within participants on average by *r*=.04 (*SD*=0.11), we think that this effect is not simply be due to a speed-accuracy tradeoff.

The fact that the route direction effect disappeared for later pointings further supports the idea that later pointings build upon earlier pointings. If for every target a new incremental construction process was initiated, we should have observed a similar route direction effect as in the first pointings. Subsequent construction from the previous target was equally often along as well as against route direction: as participants pointed in target sequences towards and away from them later pointings always incorporated both route directions and any difference would average out. Therefore, no route direction effect would be expected, just as was observed in our experiment[14].

Limits and Alternative Explanations

Our results are well explained by forward directed graph models. They account for the observed effects of route direction and distance on error and latency when performing the first trial within a chunk of related trials and can explain the absence of these effects for later trials. However, there are some alternative explanations and considerations that need to be addressed before getting to the conclusion.

The process of recalling survey estimates from a directed graph was described before by Meilinger (see introduction): constructing a mental model of the surrounding non-visible space [7]. Alternatively, navigators could *mentally walk* through a fully integrated cognitive map following the path they walked during learning. While mentally moving from one point to another, they use their path integration system to integrate the metric survey relation between their starting location and their mental position in the map until reaching the target [4, 21], resulting in a homing vector pointing back to their actual, current location. By inverting the resulting vector survey estimates from the location to the target can be derived. The activation pattern of hippocampal place cells is a plausible mediator for this process, although the conscious imagery of the mental walk might take place in posterior parietal cortex. Place cells represent locations within an environment. Even in the absence of sensory stimu-

---

[14] If later estimates were based on estimates of previous targets, the route direction effect for later pointings should invert in the case of towards pointing (see examples in Fig. 2, right). Initially the most distant location must be constructed followed by closer targets, hence, moving along the graph structure in the opposite direction compared to the first target. This inversion for towards pointings is not reflected in participants performance (see footnote 1, no meaningful interactions with target sequence). Thus, the route direction effect does not seem to change in a meaningful way as a function of target sequence. Here participants might have also accessed previously constructed mental model parts still present in working memory. The role of route direction for later pointing thus is not yet fully clear.

lation (e.g., during sleep) they can fire in an ordered fashion as they would do when walking a route [22] and such firing patterns were shown even when stationary within an environment [23]. Similar neural processes might happen during mental walks when performing a survey task. Such a mental walk process is also constructive and incremental but not based on a single graph structure.

Importantly, our findings regarding the route direction effect exclude the possibility that pointing relied exclusively on a cognitive map that abstracted from the walked direction, for example, a coordinate system. Assuming a process of mentally walking within a fully integrated cognitive map is not sufficient to explain the observed route direction effect. However, it is possible to account for this effect if survey relations are stored in such a map layer in addition to an asymmetric route knowledge layer [6, 16]. This route layer then must be involved in generating the survey estimates to introduce the observed asymmetries based on the mental walk approach.

Our study extends findings from [14] where participants learned a route in both directions and asymmetries in pointing accuracy were observed. These results could have relied on two separate and differently distorted maps for each walking direction. This is no viable explanation for the result of the present experiment. The learning experience in our study was uni-directional and effects were found both in error and latency. Thus, the asymmetry must be intrinsic to the memory of a single walked direction. Overall, the route direction effect shows that the long-term memory used for pointing must be oriented – either in form of a directed graph or a combination of a directed route layer and an undirected survey layer.

There is an important aspect inherit in the interpretation of our results of route direction as forward encoding, namely, the integration from one's current location towards the target. Such an "away" integration is assumed by both constructive positions, the mental model and the mental walk. However, our data can also be explained otherwise, namely by reversing the assumptions of forward encoding and integration away from the current location into route backward encoding and integration from the target towards the current location. While no theoretic position clearly proposes this possibility, it is still a conceivable alternative explanation that should be considered and discussed. For their first pointing participants might imagine themselves standing at the target location, mentally walk from there towards their current location while updating the vector towards the target. The resulting vector points towards their target. Importantly, to point correctly participants then must align the orientation when mentally arriving at their current location with their actual, physical orientation at that location as both will differ in most of the cases. For a backwards encoded route this process is quicker and/or more accurately for targets located towards the end (i.e., mentally walking route backwards to the current location) than for targets located towards the start when mentally walking route forwards opposite to encoding. Such a backwards route encoding might be based on spatial updating of previous locations while walking to the next location during learning, thus resulting in vectors pointing backwards. Potentially, navigators then could update not just the last visited intersection, but all previously visited locations as proposed by Wang [24]. For later pointings the previous target vector from current to the first target location first has to be inverted again, the navigator mentally teleported to the old target location which again involves an alignment of the current orientation and the mental orientation taken at the old target. Only then vector updating while mentally walking from the old target to the adjacent novel target can start. While not impossible, the required vector inversions with their associated alignment processes do seem cognitive demanding.[15]

Overall, the reverse model based on the assumption of backwards route direction encoding and integration from the target towards the current location is consistent with our data. Yet, it is disconnected with other theoretic positions, it requires the assumption of cognitive demanding inversion processes, and it is not able to incorporate findings from the literature that clearly support forward encoding. For example, the route direction effect in landmark recognition [18–20]. Furthermore, recognition triggered response models for route knowledge [6, 16] and supporting evidence from route choice [25] also are intrinsically forward oriented. Support for the mental path integration away from ones current location towards the target is given by successive activation of hippocampal place cells along a path to the goal [23]. We think that the easiest explanation and most consistent with the literature is that the route was encoded in walking forward direction and participants integrated from their current location towards the target either by constructing a mental model of the non-visible surrounding [7] or by mentally walking there and using path integration to estimate the resultant vector based on a cognitive map [4, 21] and an additional layer of directed route knowledge.

[15] Note that the mental walk model faces similar inversion problems for later pointings. No such inversions are required when pointers construct a mental model of their non-visible surrounding based on a graph representation from their current location towards the target which then is mentally “visible” as an ego-centric vector.

We are confident that our study provides a reliable basis for our conclusions. Participants learned a highly controlled but realistic city environment and learned from physically walking real life distances on an omnidirectional treadmill involving proprioceptive and vestibular cues. While the sample is not too large (23) it comprises of roughly 50% females and males and spans from 21 to 64 years of age showing a comparatively broad age spread. Furthermore, over 7200 data points went into our analysis which minimizes any random effects. The different comparisons and parameters such as route direction and distance effects across first vs. later pointings nicely correspond and are theoretically and empirically well connected.

We clearly cannot exclude that direction estimates sometimes relied on strategies rejected here. However, based on the strengths mentioned we think that such strategies can only comprise in a small minority of trials or persons in the present data. For generalization to other situations it is clear that different learning situations can result in different representations and estimation processes such as learning from maps vs. navigation [26, 27]. The reduced visual field and the instruction to not look back towards where participants came from during learning slightly limits generalizability of results as this restriction partly prevents natural navigation behavior. However, support for asymmetries in spatial memory were found in survey estimates despite learning the environment in route-forward and -backward direction [14]. Overall, we belief that our findings apply to real live-experiences when navigators learn a large-scale space exclusively from navigation. Based on our results we cannot exclude the possibility that global integration into a cognitive map and full abstraction from the directedness and incrementality of the learning experience might occur, for example, with extensive exposure to a sufficiently small environment. Nevertheless, one of the main insights from our study remains: to be able to make survey estimates in navigable space it is not necessary to rely on a globally consistent cognitive map. Survey estimates can and seem to be generally based on piecewise spatial knowledge connected by directed links that is used to incrementally recall target locations on the fly.

Conclusions

The most plausible interpretation of the present results in the light of previous findings and theoretic considerations is that participants encoded the environment route piecewise in route-forward orientation and integrated this information incrementally during survey estimates from their current location towards the first target and from there onwards to later targets. Following the mental model approach, this estimation process relied on a directed graph memory of the space. When extending the mental walk approach, it can likewise explain the results by assuming that the direction estimation is based on a combination of a directed route layer and an undirected survey layer (cognitive map). Importantly, we showed that later pointings depended on earlier pointings. Overall, our results add to the growing evidence that survey estimates obtained via navigation are constructed incrementally during recall and they further show that also survey knowledge is intrinsically oriented.

Acknowledgments

This research was supported by the DFG grant ME 3476/2-1. We thank Jan Souman for help in planning the experiment and discussing the results, Nadine Simon for help in data collection, as well as Joachim Tesch, Michael Kerger, and Harald Teufel for intensive technical support.

References

[1] Mou, W., McNamara, T.P., Valiquette, C.M., Rump, B.: Allocentric and egocentric updating of spatial memories. J. Exp. Psychol. Learn. Mem. Cogn. 30, 142–157 (2004).

[2] O'Keefe, J.: An allocentric spatial model for the hippocampal cognitive map. Hippocampus. 1, 230–5 (1991).

[3] Gallistel, C.R.: The organization of learning. MIT Press, Cambridge, MA (1990).

[4] Byrne, P., Becker, S., Burgess, N.: Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychol. Rev. 114, 340–75 (2007).

[5] Chrastil, E.R., Warren, W.H.: From cognitive maps to cognitive graphs. PLoS One. 9, e112544 (2014).

[6] Mallot, H.A., Basten, K.: Embodied spatial cognition: Biological and artificial systems. Image Vis. Comput. 27, 1658–1670 (2009).

[7] Meilinger, T.: The network of reference frames theory: A synthesis of graphs and cognitive maps. (2008).

[8] Fujita, N., Klatzky, R.L., Loomis, J.M., Golledge, R.G.: The Encoding-Error Model of Pathway Completion without Vision. Geogr. Anal. 25, 295–314 (1993).

[9] Meilinger, T., Riecke, B.E., Bülthoff, H.H.: Local and global reference frames for environmental spaces. Q. J. Exp. Psychol. 67, (2014).

[10] Werner, S., Schmidt, K.: Environmental reference systems for large-scale spaces. Spat. Cogn. Comput. 1, 447–473 (1999).

[11] Strickrodt, M., Bülthoff, H.H., Meilinger, T.: Memory for navigable space is flexible and not restricted to exclusive local or global memory units. J. Exp. Psychol. Learn. Mem. Cogn (in press).

[12] Marchette, S.A., Ryan, J., Epstein, R.A.: Schematic representations of local environmental space guide goal-directed navigation. Cognition. 158, 68–80 (2017).

[13] Meilinger, T., Strickrodt, M., Bülthoff, H.H.: Qualitative differences between environmental and vista space memory are caused by the separation of space, not by movement or successive perception. Cognition. 155, 77–95 (2016).

[14] Meilinger, T., Henson, A., Rebane, J., Bülthoff, H.H., Mallot, H.A.: Humans Construct Survey Estimates on the Fly from a Compartmentalised Representation of the Navigated Environment. In: Spatial Cognition XI (same volume).

[15] Wilson, P.N., Wilson, D.A., Griffiths, L., Fox, S.: First-perspective spatial alignment effects from real-world exploration. Mem. Cognit. 35, 1432–44 (2007).

[16] Poucet, B.: Spatial cognitive maps in animals: new hypotheses on their structure and neural mechanisms. Psychol. Rev. 100, 163–182 (1993).

[17] Stern, E., Leiser, D.: Levels of Spatial Knowledge and Urban Travel Modeling. Geogr. Anal. 20, 140–155 (1988).

[18] Schweizer, K., Herrmann, T., Janzen, G., Katz, S.: The route direction effect and its constraints. In: Freksa, C., Habel, C., and Wender, K.F. (eds.) Spatial cognition. pp. 19–38. Springer, Berlin (1998).

[19] Janzen, G.: Memory for object location and route direction in virtual large-scale space. Q. J. Exp. Psychol. 59, 493–508 (2006).

[20] Schinazi, V.R., Epstein, R.A.: Neural correlates of real-world route learning. Neuroimage. 53, 725–735 (2010).

[21] McNaughton, B.L., Battaglia, F.P., Jensen, O., Moser, E.I., Moser, M.B.: Path integration and the neural basis of the "cognitive map." Nat Rev Neurosci. 7, 663–678 (2006).

[22] Skaggs, W.E., McNaughton, B.L.: Replay of Neuronal Firing Sequences in Rat Hippocampus During Sleep Following Spatial Experience. Science. 271, 1870–1873 (1996).

[23] Pfeiffer, B.E., Foster, D.J.: Hippocampal place-cell sequences depict future paths to remembered goals. Nature. 497, 74–9 (2013).

[24] Wang, R.F.: Building a cognitive map by assembling multiple path integration systems. Psychon. Bull. Rev. 23, 692–702 (2016).

[25] Gillner, S., Weiss, A.M., Mallot, H.A: Visual homing in the absence of feature-based landmark information. Cognition. 109, 105–22 (2008).

[26] Thorndyke, P.W., Hayes-Roth, B.: Differences in spatial knowledge acquired from maps and navigation. Cogn. Psychol. 14, 560–589 (1982).

[27] Meilinger, T., Frankenstein, J., Watanabe, K., Bülthoff, H.H., Hölscher, C.: Reference frames in learning from maps and navigation. Psychol. Res. 79, 1000–1008 (2015).

7.3 Study 3: Learning a non-Euclidean environment

Study reference

Title: Navigators learn a local graph, not a global map, of a complex environment

Authors: Marianne Strickrodt, Tobias Meilinger, Heinrich H. Bülthoff, William H. Warren

Status: In preparation

Abstract

We are navigating in a complex world, yet we are able to grasp the spatial relations between the different places we visit. It is still an unresolved issue how we represent navigable space and use this information for shortcutting and survey estimates such as pointing to distant, non-visible landmarks. In this study we set out to compare two competing theoretical approaches on how navigable space is represented. We contrast the idea of the metric embedding of the spatial information we gather into a single global reference frame (i.e., Euclidean mental map) with models assuming that only local place-to-place metrics are stored that can be used to compute survey estimates when needed (i.e., labelled graph representation). Two groups of participants learned either of two circular, multi-corridor, virtual mazes containing seven objects by walking. One maze architecture was the impossible non-Euclidean version of the possible Euclidean maze. In the impossible environment the local place-to-place metrics after walking one lap would not match on a global metric level. Instead, participants were covertly teleported into the start corridor again despite being located distant from its initial Euclidean location. In a subsequent pointing task targets were queried in predefined sequences, either clockwise or counterclockwise around the circular connectivity of the environment relative to one's current location. In contrast to the possible maze group estimates of the impossible maze group violated the metric postulates of a Euclidean map: Participants did not point the same direction albeit being queried the same target. Instead their estimates were biased by the order of target sequence and strongly followed the pre-activated local place-to-place metrics. Our results suggest that navigable space is not stored in a globally consistent format such as the Euclidean map, but instead that a labelled graph was underlying survey estimates.

Keywords: spatial memory; graph knowledge; labelled graph; network of reference frames; Euclidean map; cognitive map; impossible environments

Introduction

In our everyday life we pass through and by a number of different places. For example, for doing grocery shopping we have to leave our apartment, walk a few (or more) streets, thereby maybe passing our hair dresser. On our way back from the supermarket, we might decide to take a detour to stop by at a friends' house for a coffee and head back home afterwards. One characteristic of such navigable environments is, that they can never be grasped from a single standpoint but must be experienced successively. Still we are capable to understand the relative location of the places visited and form survey knowledge of our neighborhood that allows us to estimate straight line direction and distance and to do novel shortcuts across unexplored terrain between learned places.

There are different understandings of how such survey knowledge is represented in our brain. One influential proposition is that of the cognitive map. The term was first expressed by Tolman (1948). Even though he did not necessarily want his "cognitive map" term to be interpreted in its literal sense, the term was adopted and expanded by other researchers in this vein (e.g., Byrne, Becker, & Burgess, 2007; Gallistel, 1990; O'Keefe & Nadel, 1978). In its rigid definition the cognitive map approach suggests that we form a metric, map-like mental representation of our surrounding space. The discovery of (among others) place cells and grid cells in rats (e.g., Hafting, Fyhn, Molden, Moser, & Moser, 2005; O'Keefe & Nadel, 1978) and later also in humans (in the form of intracranial recordings or more indirectly by corresponding fMRI BOLD patterns) (e.g., Doeller, Barry, & Burgess, 2010; Ekstrom et al., 2003; Jacobs et al., 2013; Jacobs, Kahana, Ekstrom, Mollison, & Fried, 2010) are often brought forward as the neurological fundament for the embedding of place information into a stable, cognitive reference system. A key component of the mental map approach is, that all places experienced can be assigned distinct places in the cognitive reference system, just as coordinates in a x-y coordinate system. An example is given in Figure 1 panel A and B, where four places (J, K, L and M), which are experienced successively in a circular manner in the external world, are represented in the form of a Euclidean map. Taken literally such a global, all-encompassing, Euclidean map must be bound to the metric postulates of positivity, symmetry, and triangle inequality (e.g., Beals, Krantz, & Tversky, 1968; McNamara & Diwadkar, 1997; Warren, Rothman, Schnapp, & Ericson, 2017). Positivity refers to the idea that the distance between any point and itself must be zero, as there can only be one point in the Euclidean mental map representing a place in the external world. Further, the distance between any two points must be larger than zero, as representation of distinct places cannot overlap. Symmetry of points is achieved when the distance estimated from point A to B is the same as the distance estimated from point B to A and when the direction vector from A to B is the exact reverse from B to A. Triangle ine

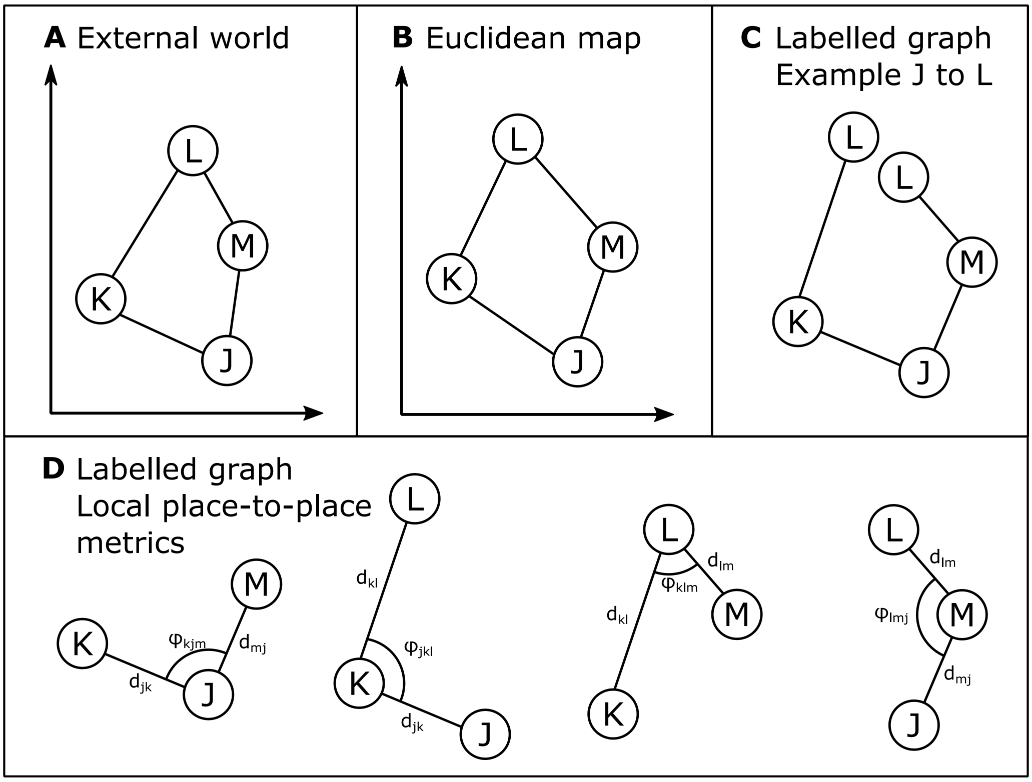


Figure 1. Contrasting predictions for Euclidean map and labelled graph representation. **A|** Example of navigable space, four places (J, K, L M) connected successively. **B|** Distorted but globally consistent Euclidean map representation of panel A. Each place is assigned to distinct coordinates which can be read-out from the coordinate system to make survey estimates. **C|** Example of recalling the location of place L when standing at place J based on a labelled graph (visualized in detail in panel D). Depending on which local place-to-place metrics are used (clockwise or counterclockwise) different estimates of the location of place L are possible. **D|** Labelled graph representation where represented translational and rotational metrics are restricted to the direct neighbors of a place. These local place-to-place metrics do not need to be globally consistent, leading to violation of metric postulates when used for survey estimates of distant places (see panel C). Note that, albeit redundant, some distances here are visualized twice for clarification reasons to grasp the idea of local metrics to direct neighbors which are not brought in global consistency.

quality defines the relationship between any three points in the Euclidean map and inherits the rules that the distance between point A and B and from there adding the distance between point B and C must always be larger or equal to the distance between A and C. Further, when assuming a Euclidean representation the inner angles of such a triangle should sum up to 180°. This can also be extended to the relationship between the four points in the Euclidean mental map in panel B that must form an irregular quadrilateral with a sum of inner angles of 360°. Such a Euclidean mental map must not be perfect in representing the external world, but translational and rotational metrics can be distorted within the limits of the metric postulates. We know, for example, that the human path integration used for judging walked distance and directions to a starting location suffers from a rather low resolution and discontinuities (e.g., Loomis et al., 1993; Zhao & Warren, 2015a, 2015b). Further, humans tend to remember irregular environments and junctions as more regular (e.g., orthogonal streets) as they are (e.g., Byrne, 1979; Moar & Bower, 1983; Tversky, 1981) and allow boundaries to bias their distances judgement between places (e.g., Kosslyn, Pick, & Fariello, 1974; McNamara, 1986; Newcombe & Liben, 1982). None of these findings contradict the idea of a Euclidean map representation per se, which can well be distorted as long as clear locations within the Euclidean mental map are assigned.

Alternative approaches emphasize that there is no need for a global metric embedding of all places in spatial long-term memory to be able to perform survey tasks such as distance and direction estimation to non-visible landmarks or novel shortcutting. Instead, what may be stored are spatial information about the connectivity of all places (topological graph with neighbor to neighbor connections) which are enriched by local information about metric rotation and translation between the neighbors (e.g., Chrastil & Warren, 2014; Meilinger, 2008; Warren, Rothman, Schnapp, & Ericson, 2017). Albeit differing in small details about the nature of local place representation and attached labels such theories can be summarized under the term labelled graph theories. Figure 1, panel *d* visualizes the idea. For example, when memorizing the connectivity of J-K-L we additional represent the distance between K and M as well as the distance between K and J. Furthermore, angular information about the direction of J relative to L when standing at K are stored. Such local place-to-place metrics that are stored for the other neighboring places as well. To solve survey tasks, for example standing at J pointing to L, one's current position and the target first must be brought into direct reference by successively recalling the stored local metrics from one's current position to the target to come up with an estimate. This can be done by vector addition (Warren et al., 2017), for example by mental imagination of how the nonvisible parts of the environment could be strung together successively (Meilinger, 2008). Like the Euclidean map representation the labelled graph considers potential distortions of those local metrics that arise during the encoding process. The crucial difference between the two approaches, however, is that these local metrics to not need to be globally consistent. An example for estimating the direction from J to L based on the local graph information is given in Figure 1, panel C. Depending on whether one mentally constructs the estimate from J to L via K (clockwise) or via M (counterclockwise) different estimates for the location of L are formed, as one follows the local metrics along either direction which each underly distinct distortions. Such distortions might be rather small and remain unnoticed in the majority of studies due to the very high absolute angular error of 20-100° which are typically found in pointing performance in navigable space (e.g., Chrastil & Warren, 2013; Foo, Warren, Duchon, & Tarr, 2005; Ishikawa & Montello, 2006; Meilinger, Riecke, & Bülthoff, 2014; Schinazi, Nardi, Newcombe, Shipley, & Epstein, 2013; Weisberg, Schinazi, Newcombe, Shipley, & Epstein, 2014). As a consequence of distorted local place-to-place metrics the metric postulates of positivity and triangle inequality can be violated on the global scale. There is not one distinct location for each place and calculating the sum of the inner angles between the four places does not need to correspond to what would be expected in Euclidean terms (i.e., not 360° in our scenario in Figure 1).

There is ample evidence for the formation of distinct memory units of individual local environments. For example, the updating of object locations was found to be impaired when leaving the test room (Wang & Brockmole, 2003a). Likewise, the updating of objects outside the test room was impaired when reorienting with respect to the test room compared to reorienting with respect to the outside of the test room (i.e., university campus) (Wang & Brockmole, 2003b). Also, knowing the exact location of an object within a vista space does not necessarily seem to be accompanied with knowledge about where this vista space itself is located (Marchette, Ryan, & Epstein, 2017). Furthermore, it was found that participants direction estimation performance in multi-corridor spaces is facilitated when their body is re-aligned with the view first experienced within each corridor the task was performed in (i.e., when looking along the corridor), compared to being aligned otherwise (e.g., looking against the wall) (e.g., Meilinger et al., 2014; Strickrodt, Bülthoff, & Meilinger, 2018). This is typically interpreted as the formation of local reference frames, mental reference systems confined to the locally visible environment (i.e., an individual corridor in the cited studies). Relative to this local reference frame locations of objects within that local environment are stored. Being aligned with the reference axis of the mental reference system alleviates the recall of stored spatial information, while being non-aligned involves effortful mental transformations to map the orientation of the reference system onto one's current perspective (e.g., McNamara, Sluzenski, & Rump, 2008; Mou, McNamara, Valiquette, &

Rump, 2004). The findings of local reference frames are often accompanied by increasing pointing latencies across increasing number of traversed corridors between one's current location and the target one has to point to (e.g., Meilinger, Strickrodt, & Bülthoff, 2016; Tobias Meilinger, Strickrodt, & Bülthoff, 2018; Strickrodt et al., 2018). Such corridor distance effects seem to reflect nicely the recall process expected for labelled graph representations. Because individual corridors (and objects within) do not seem to be represented within a single reference system, but instead the local reference frames (i.e., memory units) had to be successively activated along the order of experience until reaching the target.

Previous studies already found evidence contradicting the metric postulate of symmetry for Euclidean mental maps. It was shown, that a number of spatial performance measures differed between the same pair of places, depending on whether they are made from place A to B or from place B to A. Such asymmetries were found in route selection (Stern & Leiser, 2010) and distance estimations (Burroughs & Sadalla, 1979; McNamara & Diwadkar, 1997; Sadalla, Burroughs, & Staplin, 1980). They indicate that the underlying memory structure might not be a Euclidean representation. Over the years, however, alternative models and processes were suggested that could account for such asymmetries while leaving the assumption of a Euclidean mental map with symmetric distance representation untouched. Two bias models have been introduced that define the cause for asymmetries not in the representation of spatial properties but in biases introduced by the estimation process itself. The category-adjustment model of spatial coding (Huttenlocher, Hedges, & Duncan, 1991; N Newcombe, Huttenlocher, Sandberg, Lie, & Johnson, 1999) assumes a hierarchical representation on a fine-grained (conceivably Euclidean) and a categorical level. Biases are produced during retrieval by adjustments made to the fine-grained place information based on its distance to a category prototype. A typical category prototype is the middle of a quadrant of a circle which is drawn on a piece of paper and formed by vertical and horizontal visual axes. Biases in location estimations towards the center of category are expected, resulting in asymmetric biases. In contrast, the contextual-scaling model (McNamara & Diwadkar, 1997) assumes that different places evoke different contexts in working memory when being referenced depending on their salience (e.g., familiarity, functional importance). Thus, a different context is activated when standing at place A recalling the location of B compared to standing at place B recalling A. This scales the retrieval process accordingly and leads to asymmetries. Both models allow for the existence of a

Euclidean mental map and renders the asymmetries found in route choice and distance estimation much less of a violation of metric postulates than originally thought. Along this line, it was shown in a few experiments that besides local reference frames for each visited place also global reference frames can be formed and used by participants. The self-localization time (Meilinger et al., 2013) as well as the direction estimation process (e.g., Strickrodt et al., 2018; Tlauka, Carter, Mahlberg, & Wilson, 2011; Wilson & Wildbur, 2004; Wilson, Wilson, Griffiths, & Fox, 2007) within navigable space was shown to be facilitated when participants were oriented along a uniform global orientation during testing, independent on where in the environment estimates were made. Such an environment-encompassing orientation dependent recall along a global orientation can be interpreted as the formation of a reference system that is covering and inheriting place information from the entire environment learned, thus, being an indicator for a global metric embedding.

In our study we took a rather novel approach to examine the Euclidean nature of spatial representation. We investigated how the human spatial memory system deals and makes sense of non-Euclidean, impossible environments that are learned in virtual reality. As a baseline we constructed a normal, possible virtual environment and then manipulated the architectural properties such as angles between connecting corridors or length of corridors to produce an impossible version of it (see Method). This is one technique which is occasionally used in the virtual reality community to enable users to experience a large virtual environment despite walking in a limited physical space. There, often overlaps of virtual rooms that are visited successively are produced, with the aim to be imperceptible to the user (e.g., Suma, Lipps, Finkelstein, Krum, & Bolas, 2012; Vasylevska, Kaufmann, Bolas, & Suma, 2013). Another technique is the manipulation of the mapping of virtual and real rotational and translational gain. Both result in what the VR community calls “redirected walking” (see Nilsson et al., 2018 for a review). In our study we concentrate on the first technique while leaving the real and virtual gain perfectly matched and untouched.

Impossible architectures have been used before in the spatial cognition literature. Zetzsche, Wolter, Galbraith and Schill (2009) showed that participants could successfully find the shortest routes to objects within impossible environments learned from a 2D projection (see also Ruddle, Howes, Payne, & Jones, 2000; replicated in walkable environment by Warren et al., 2017). Kluss, Marsh, Zetzsche and Schill (2015) used a complex virtual reality setup where participants wore a head-mounted display and walked on a treadmill to explore simple environments. For example, one possible environment consisted of three corridors forming a triangle (sum of inner angle 180°) whereas the impossible version imitated walking three legs of a triangle albeit 90° connecting angles (sum of inner angles 270°). Participants subsequent task was to reproduce one walk through each environment blindfolded. Examining participants turning angles showed that in sum they were comparable to what would be expected when relying on local metrics in the impossible maze (sum of angles around 270°). However, due to a high variability in responses the authors fail to show that this angle sum is different from the behavior found in the possible triangle environment (180°). Therefore, albeit indicating that local metric knowledge was acquired without detecting that an impossible maze was walked, these results cannot exclude that participants adjusted the local metrics experienced during learning to match up globally and fit into a Euclidean mental map.

In a study by Warren and colleagues (2017), participants learned the location of eight objects within a complex, 11m x 11m virtual hedge maze by walking in a large tracking hall being equipped with a head-mounted display. For one group of subjects they installed virtual "wormholes" that, upon contact, rotated the maze by 90° around the midpoint without participants noticing any visual change in the current corridor view. Thus, for some of the objects the local metrics from object A to object B when walking there via the constant midpoint of the maze would be globally inconsistent with the local metrics traversed when walking back from B to A but via a route crossing a wormhole. When participants were instructed to perform straight line estimates between the learned objects strong distortions towards the wormhole locations were found for objects near wormhole entrances. Compared to the possible maze group where no rotation occurred, near-wormhole locations in the impossible maze were represented as ripped apart from other close-by object locations thereby folding over to be represented closer to other objects actually farther distant in the visual reference frame of the environment. It remains somewhat unclear whether those biases in direction estimates can also be explained by a highly distorted but overall globally consistent Euclidean map, because it is hard to make clear, testable predictions for a potential global embedding. Indeed, using a different wormhole maze Muryy and Glennerster (2018) found that participants pointing behavior could be well modelled by assuming distortions to the locations and orientations of local places. However, the study by Warren and colleagues (2017) clearly shows how local metrics experienced from one object to the next strongly shape and influence the (potential graph) structure of the representation. If global embedding is attempted, it clearly involves the integration and adaption of multiple local metric information.

How might global embedding be achieved? It is generally known that humans can update the position of landmarks in their environment without visually attending them (e.g., Farrell & Robertson, 2000; Martin & Thomson, 1998; Wang & Spelke, 2000). For this updating process path integration provides useful information. By introducing sudden conflicts in homing tasks, it has been shown that humans can integrate information acquired from self-motion cues gathered via path integration with visual landmark information that were covertly changed on the return path (e.g., Cheng, Shettleworth, Huttenlocher, & Rieser, 2007; Nardini, Jones, Bedford, & Braddick, 2008; Zhao & Warren, 2015). We know from research with desert ants that the path integrator can be reset at known location (i.e., the nest) to remove error that accumulated during foraging (e.g., Müller & Wehner, 1994; Wehner & Srinivasan, 1981). The gain factor of the path integration system itself was shown to be modifiable when introducing a mismatch between visual and idiothetic translation and/or rotation gain during normal exploration of an environment both in rats (Jayakumar et al., submitted) and men (Tcheang, Bulthoff, & Burgess, 2011). These studies suggest a strong coupling between visual landmark cues and idiothetic self-motion cues to make sense of the space around us on a global level. In a relatively new theoretic approach Wang (2016) proposes the use of multiple path integrators to simultaneously track and update the location of multiple targets during walking a track as basis to form a cognitive map of this environment. Vectors to all the object and places of interest are carried along the way which are all subject to possible recalibrations when in view of a known landmark. If such updating and recalibration processes take place in more complex, navigable environments as well they constitute a mechanism that supports Euclidean embedding of impossible architectures. In contrast to achieving global embedding during the encoding process it could also be accomplished by performing local triangulation on the long-term memory of stored metrics. Like this, inconsistencies can be removed and the representation globally optimized (e.g., Mallot & Basten, 2009).

In our study we set out to test the two models for the representation of human survey knowledge by contrasting clear predictions for the reliance on either local place-to-place metrics that were experienced during learning (labelled graph) or reliance on a globally embedded Euclidean mental map. Two groups of participants learned either of two circular, multi-corridor, virtual mazes, one being the impossible, non-Euclidean, torn-apart version of the possible maze. They then had to estimate the straight-line direction between pairs of objects within the environment. During this testing phase the order targets were queried was manipulated, following either the clockwise or counterclockwise object sequence around the circular connectivity of the environment. If a labelled graph representation is underlying the estimation process, direction estimates should be based only on the local place-to-place metrics that are activated by the biased target sequence (compare to Figure 1, panel C). In the impossible maze case this procedure should lead to clearly different direction estimations for one and the same reference-target pair because of the impossible architecture that was learned. In contrast, if global embedding takes place this should involve distorting the perceived local place-to-place metrics to form a Euclidean map obeying the metric postulates. The questions we asked are: Do participants in the impossible maze group always point the same direction when queried the same object (positivity postulate)? And how much do their estimates either fit to the directions predicted when following local place-to-place metrics or a global embedding?

Method

Participants

Forty-eight participants with normal or corrected to normal visual acuity partook in the experiment, all naïve to the research question and the experiment, receiving monetary compensation for participation. Without their knowledge 25 of them were assigned to learn the possible maze, 23 to learn the impossible maze. The experiment was terminated before completion for two participants, one in each maze type group. One requested to stop early during the testing phase as he found the task too hard. For the other one the learning phase could not be completed because of equipment failure. In the possible maze group another six participants and in the impossible maze group another four had to be excluded from the analysis because of chance level performance in the testing phase (see Results). Eighteen participants remained in the possible (M 7, F 11, age M=29.61, SD=9.73) and 18 participants in the impossible maze group (M 5, F 13, age M=28.94, SD=12.46) that were included in the analysis. Half of the participants in each maze type group (nine each) started learning the environment walking clockwise first, the other half walking counterclockwise first (see Procedure). The experiment was approved by the local ethics committee.

Material

The experiment took place in a large tracking space of 12×12 m equipped with 20 high-speed infrared cameras (Vicon® MX 13). The system tracked participants' head position at 180 Hz and transmitted the information wirelessly to a portable computer (MSI VR One) carried by the participant. The egocentric view within the virtual environment was displayed to the participants via a head-mounted display (HMD), an Oculus Consumer Version 1 (CV1), rendered with a NVIDIA GTX 1070 graphics card. The CV1 provided a field of view of ca. 110° diagonal at a resolution of 1080x1200 pixels for each eye. This immersive virtual reality setup involved visual as well as idiothetic cues and allowed a self-determined exploration of the virtual world, thus, delivering a realistic learning experience.

The following measures were taken to prevent participants from using real-world reference points to anchor their spatial knowledge: After being equipped with the portable backpack and the HMD participants were disoriented before learning the virtual environment by being led on a random path through the tracking hall while the HMD rendered a black screen, thereby loosing track of their orientation and starting position in the tracking hall. The experimenter followed participants closely during learning to eliminate directional auditory cues from an otherwise constant position when giving further instructions (and to ensure safe movement through the tracking space). Additionally, bird sounds were played to the participants' via HMD headphones throughout the whole learning and testing phase to prevent any audio cues (e.g., conversation from outside the tracking hall, sounds from slamming doors or construction works) to be used as a reference direction.

We ran two groups of participants, each group learning either of two virtual environments, a possible or an impossible, non-Euclidean maze. Both environments consisted of seven interconnected corridors varying in corridor length and connecting angles, thereby forming a complex, circular environment (see Figure 2). This means, after passing through the seventh corridor participants would end up in the corridor that was experienced first again. The circularity of the environments enabled participants to continuously travel through the environment for multiple laps (clockwise and counterclockwise). The corridors all had a corridor width of 90cm and each corridor contained a picture of an object hanging about eye height centered on the wall of a half-round alcove. In each corridor the alcoves were located at the inner walls with respect to the circular environment. Thus, when traversing through the corridors alcoves would always be located on participants' right when walking clockwise, but located on participants' left when walking counterclockwise.

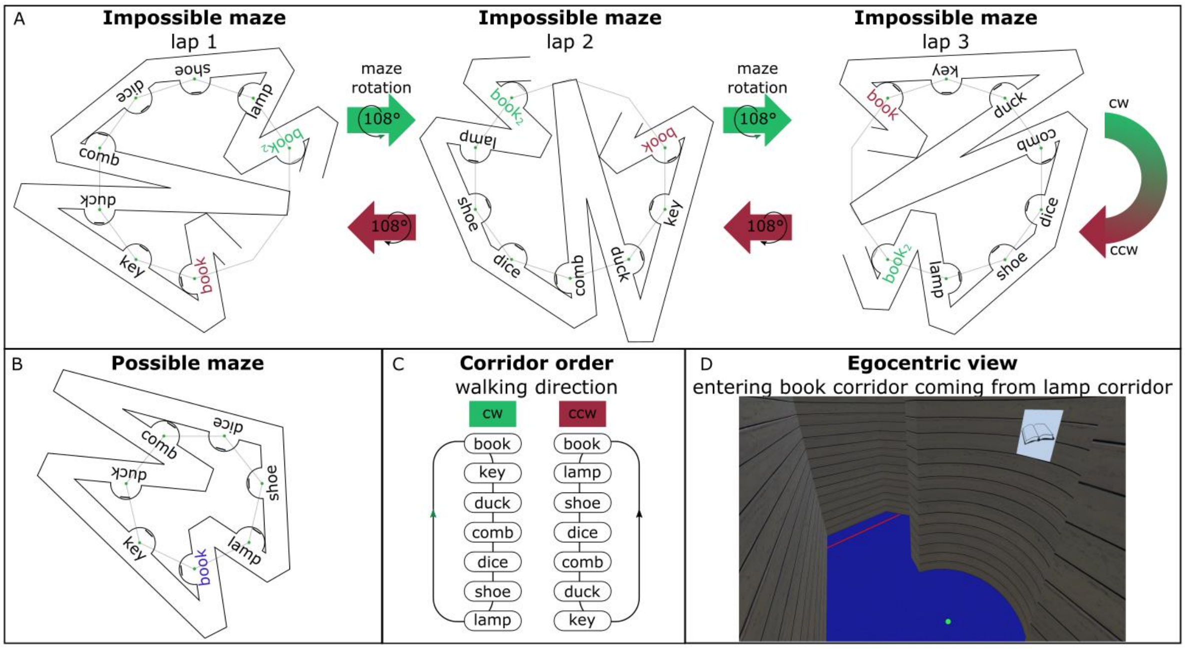


Figure 2. The two virtual environments that had to be learned and the procedure of the learning phase. **A+B|** Birds eye perspective of the mazes. Corridors are connected to form a complex environment with left and right non-orthogonal turns. The seven places to be learned (marked by green dots) are evenly spread across seven corners of a decagon (impossible maze) or a heptagon (possible maze), visualized here in grey. Within each corridor each place is identified by an object hanging in an alcove (small black lines at center of alcove wall). Participants could fully look around when standing on a green dot. **A|** Layout of impossible maze and example of learning procedure when starting clockwise (cw). Lap 1 is started at the book corridor (marked here in red) and continued clockwise (key, duck, etc.), until reaching a book corridor identical to the first one (marked here in green, $book_2$) without having reached the original location in physical space. At this point the virtual maze was rotated by 108° (without participants noticing) to match up the first book corridor with participants current position in the maze. Visually seamless lap 2 started (middle of panel A) and participants continued walking clockwise till reaching the book corridor distant from where they started off again (marked in green, $book_2$). The maze was rotated again. When reaching the book corridor a third time after lap 3 ($book_2$, right side of panel A) participants changed walking direction and returned along their previous path in the counterclockwise corridor order for another three laps (panel A from right to left). Like this the local corridor-to-corridor metrics remained constant across each lap although the space was not matching up on a global scale in physical space. **B|** Layout of possible maze. Global and local metrics match up to form an Euclidean space. After one lap the same physical location of the book was reached again (marked here in blue). Except for no rotation of the environment the learning procedure was identical to the impossible maze group. **C|** Corridor/object order when learning the environments. Each participant walked the environment three times clockwise (cw) then three times counterclockwise (ccw) (or vice versa) before undergoing a first learn-check. **D|** Example of an egocentric view. A green dot on the floor indicated the exact place that had to be learned and was identified by the object hanging in the center of the alcove wall. A wooden panel structure was attached to the walls, a blue carped on the floor. Red lines at the connection between corridors mark the walkable area.

The diameter of each alcove was 60cm and their center marked by a green dot on the floor. These green dots marked the places participants had to memorize. Thus, each place was unambiguously localized by the green dot on the floor in each corridor and labelled by the respective object hanging in the alcove. The following objects were depicted in black-and-white on a 25 x 25cm plane (2D) floating in the alcoves, thereby identifying the places to be learned in the consecutive corridors: book, key, duck, comb, dice, shoe, lamp (clockwise order). We selected monosyllable objects. Identifiability was verified in a pre-test. Four naïve participants (not included in the sample of this study), three of them native English speakers, named the items correctly and unambiguously.

The possible maze served as the baseline: a complex but Euclidean space. Corridors differed in length and connecting angles, however, the places to be learned (green dots at the center of alcoves) were evenly spaced around an imaginary circle, forming the layout of a symmetric heptagon (i.e., seven-sided polygon). The straight-line distance between direct neighbors (e.g., dice and shoe, key and duck) was 2.38m, corresponding to the side length of the heptagon underlying the place layout. The interior angle (i.e., angle between two adjacent sides of the heptagon) was 128.57°. This corresponds to, for example, the pointing angle between shoe and comb when standing at the dice or the pointing angle between duck and book when standing at the key. All non-adjacent objects relative to one's current position (e.g., lamp, book, key, duck when standing at dice) were evenly spread within this interior angle defined by the two direct neighbors, following an angle difference of 25.71° when pointing from neighbor to neighbor (e.g., standing at the dice the angle difference between shoe and lamp, then between lamp and book, between book and key, key and duck, as well as angle difference between duck and comb is 25.71°). Because of the symmetry of the underlying polygon the neighbor distance, interior angle and neighbor-to-neighbor angle difference is constant across all locations. This structure makes pointing performance across all locations comparable, thus, allowing to average them for the analysis.

The impossible maze was a distorted version of the possible maze. We disjointed the connection between the first (book) and seventh corridor (lamp) and pulled apart the layout of the seven places to be spread across seven adjacent corners of a symmetric decagon (ten-sided polygon). Individual corridor length and angles between adjacent corridors were adjusted to fit this new layout (corridor width remained 90cm), thereby keeping the overall path length of one lap similar to the path length of the possible maze (about 33.3m ±5cm). Likewise, the succession of left and right turns and the straight-line distance between direct neighbors (2.38m) was kept constant across both mazes. This ensured comparable memory load. The underlying decagon had an interior angle of 144° and a neighbor-to-neighbor (corner-to-corner) angular difference of 18°. The deformation of the impossible maze determined that after the participant walked one round (leaving the seventh corridor entering the first one again) they ended up at a position in real, physical space a few meters distant from where they started off. At this point the impossible, virtual environment was rotated by 108° within milliseconds without participants noticing any change in the visual scenery, matching up the first corridor with participants' current position (see Figure 2, panel A), enabling the participant to walk another lap. Like this, the local corridor-to-corridor metrics remained constant across each lap although the maze was not matching up on a global scale.

In both maze types the connection between two corridors was always slightly longer than needed for walking the environment. Corridor walls were elongated until they could be connected by an approximately 90cm wall (not until they met – this would have led to redundantly long corridor connections and overlap with other parts of the possible maze). We decided for this visual cue as the exploration of the maze was restricted by learning rules, one of them was to not turn around to investigate the corridor one came from (read procedure further below). Like this, exploration of the connection of the corridors was slightly constrained. The elongation of corridor connection should therefore serve as a visual cue to understand how two corridors are connected to each other. The resulting impossible maze was slightly bigger than the walkable space in the tracking hall. To prevent collisions with the walls of the tracking hall we decided to restrict the walkable area within the virtual world in a constant format, placing red lines on the floor in a distance of 90cm from every sharp corner one turns around to enter the next corridor. Participants were instructed to not walk the area behind those lines. The restriction to the walkable area of the corners was assimilated in the possible maze as well to equalize the learning experience.

For the testing phase audio stimuli were recorded to guide participants through each trial. Recording, preprocessing and cutting of the audio recording was done with praat (http://www.fon.hum.uva.nl/praat/). An adult male speaker, native English, produced instructions for guiding the testing phase, namely, the self-localization ("You're at the [object].", "Do you know where you are?") and the direction estimation phase ("Face the [object]").

Procedure

The learning and testing phase were similar for both maze types with the only difference that the maze was rotated every lap during learning in the impossible maze group to enable a natural walking experience for multiple rounds.

Learning phase.

The experimental session started with a briefing about the procedure. Participants were informed that they first had to learn the layout of a virtual environment and the places within (green dot associated with an object) and that their spatial memory for that environment was about to be tested in a subsequent spatial task. They were told: “Your task is to memorize as accurately as possible the locations of these places in the environment. Later, in the test phase, in each trial you will be teleported to one place and asked to indicate the exact locations of the other places from your current position.”

After being equipped with HMD and portable backpack participants were disoriented by several random left and right turns (HMD display turned off) while being led to the position of the starting corridor, which was always the book. Then the virtual reality was presented to them. In a pseudo-random fashion, participants either started walking the environment in the clockwise (cw) or a counterclockwise (ccw) direction. This was done to prevent potential directional effects to bias the analysis. During the learning phase the experimenter gave further instructions. Participants could explore the environment for six laps, three consecutive laps cw then turning around and walk three laps ccw (or vice versa). They could walk in their own pace but were restricted to walk or look back to where they came from (except of after three laps when they had to change walking direction). This was done to circumvent participants in the impossible maze to recognize the rotation of the maze. There were exceptions to the “do not look back”-rule, namely, when participants stood right on top of a green dot in the middle of an alcove that marked a place. There they could take a full look around, visually exploring the entire corridor and the connection to the two adjacent corridors. Participants were informed that the red lines on the floor should not be passed, but elongated connections could be used as visual cues to understand the angles corridors are connected with. During their first lap participants had to name each place/object to ensure correct identification of the place labels. Participants could ask questions about the experimental procedure throughout the whole experiment, however, questions regarding the nature and structure of the environment (e.g., “How many corridors/objects are there?”) were not answered, as the environment had to be self-explored.

After six laps through the environment a learn-check was carried out. We deprived the environment of the objects in the alcoves and had participants walk another two laps cw and two laps ccw to recall from memory the names of every other place they passed. All seven corridors were queried in each walking direction. In case of errors participants again had to learn the environment for two more laps, one cw one ccw (same order as during initial learning), followed by the identical learn-check. This was repeated until participants reached 100% accuracy in identifying the places. As soon as they reached the learning criterion the learning phase was concluded, and the testing phase started after a short pause. If the learning criterion wasn't reached within a maximum of six learn-checks the experiment was terminated and the participant sent home without completing the testing phase.

Testing phase.

After a short pause of approximately five minutes and written instructions on the task participants would have to perform, they were equipped again with HMD and portable backpack and the testing phase started. The task was conducted standing at a fixed position in the tracking hall. The testing phase composed of multiple sets of coordinated trials. Refer to Figure 3, panel A and B, for examples of trial sets for the possible and impossible maze group. At the beginning of each set of trials participants were teleported randomly in virtual space to one of the seven places that had to be learned, standing on top of a green dot in the middle of the alcove. They could fully rotate their body around the spot but were instructed to not walk away from the green dot. Their view along the corridor was blocked by a white fog (see Figure 3, panel D). Thus, all they saw was the object hanging in the alcove, the alcove and the opposite wall and a few centimeters to either side of the corridor. They were not able to see the connection to the adjacent corridors. Participants had been informed about the absence of these visual cues already during the learning phase. The object in the alcove as well as the audio instructions (e.g., "You're at the shoe.") served as cues for self-localization and orientation. Participants had to press a button on a handheld controller to indicate familiarization with and self-localization at their current position (audio instructions "Do you know where you are?"). Thereupon the object in the alcove disappeared, while subjects remained at their current location in the environment, and a target object was given via audio instructions (e.g., "Face the lamp."). At the same time a vertical, black line appeared in the center of participants visual field, following every head movement the participant made. Participants task was to align their head and therefore the black line with the location of the green dot of the target place. Thus, they had to indicate the straight-line direction towards the remembered location of the target place by directly facing it "as if the walls would be made of glass" (instructions of experimenter). To do so participants could fully rotate their body and head around the spot. After confirming the

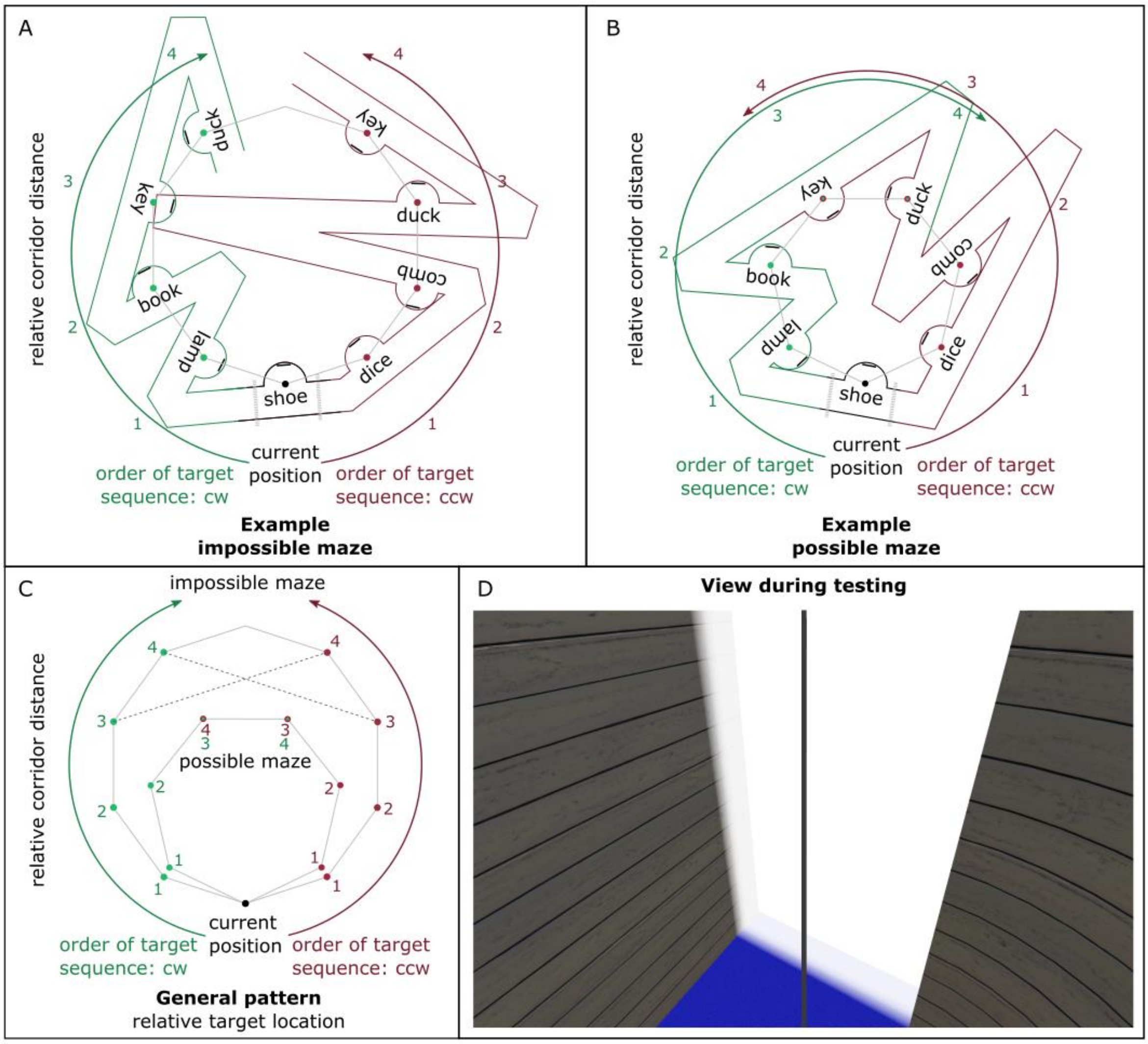


Figure 3. Testing phase. **A+B|** Two example sets of trials from the possible and the impossible maze. Standing at the shoe (black corridor) within one set of trials participants had to face four targets in a row, either along a clockwise (cw) target sequence (green corridors and arrow) with increasing relative corridor distance (1-4), or along a counterclockwise (ccw) target sequence (red corridors and arrow). Visualized are the local place-to-place metrics from corridor to corridor which had been experienced during learning. **A|** In the possible maze group we expected that participants point out the same directions to the same target whether queried in cw or ccw target sequence. **B|** For the impossible maze the local place-to-place metrics do not match up on a global scale. Depending on the order of target sequence (cw or ccw) the use of local metrics should lead to different estimated directions for the same targets. For example, the facing direction of key and duck (relative corridor distance three and four) should yield average estimates shifted to the left for the cw target sequence compared to the ccw target sequence where estimates should be shifted to the right, just as if participants would point to two different locations. **C|** General pattern of the target location based on local metrics relative to one's current location for the possible (heptagon) and impossible (decagon) maze group. This pattern can be applied to any location participants are tested from because of the symmetry and even spread of the places across heptagon or decagon. Compared to the possible maze we expect an outwards bias for the impossible maze group when following local place-to-place metrics. **D|** Example of an egocentric view during testing when looking to the left while standing at the green dot of one's current location. The view along the corridor was blocked to both sides of the alcove by a white fog. The black line at the center of participants field of view (following every head movement) should be used as an aiming device to face the straight-line direction towards the target place.

facing direction with a button press another target was given to the participant. Participants were instructed to face the target places as accurate as possible and execution was not constrained by any time limit. No feedback was given about the accuracy of their performance. At each position, participants were asked to face four targets, one by one, composing one set of trials. Then they were teleported to a different location in the environment and a new set of trials began, again starting with self-localization, followed by four target objects to face. At the beginning of the testing phase participants were familiarized with their task during four randomly chosen sets of trials. Then data collection started.

The sequence of the four targets was predefined in each set of trials. Participants either had to face four targets in consecutive order following the clockwise walking direction (cw), starting with their direct neighbor, followed by the subsequent neighbor and so on. An example is visualized in green in Figure 3, panel A and B. Standing at the shoe participants first had to face the lamp, then the book, key and finally the duck before being teleported to a new location. Alternatively, the queried target sequence followed the counterclockwise walking direction (ccw), again starting with the direct neighbor followed by the subsequent neighbors. An example is visualized in red in Figure 3, A and B: standing at the shoe participants first had to face the dice, then the comb, then duck and key. Thus, along a set of trials the corridor distance relative to one's current location increased along the predefined order (cw or ccw).

Following the assumptions made by graph theories (labelled graph or network of reference frames) recall of spatial knowledge should follow the sequence of stored local place-to-place metrics. In a circular space to access spatial information of any other place in the environment there are two possible sequences of nodes and edges one might utilize for retrieving the relative location of that place, namely, along the clockwise or the counterclockwise place-to-place sequence. Therefore, the manipulation of the retrieval direction with cw or ccw target sequences was done to bias the order in which local place-to-place information are accessed. In the case of the impossible maze those local place-to-place metrics are globally inconsistent leading to distinct predictions of the facing behavior depending on the biased direction, as depicted in Figure 3. Querying four consecutive places either clockwise or counterclockwise in an environment consisting of seven corridors involves an overlap in the last two targets of each sequence. As can be seen in Figure 3, panel A and B, standing at the shoe the two targets key and duck (with the relative corridor distance three and four), are pointed to in both cases, the clockwise and the counterclockwise order. While in the possible maze following both directions should lead to roughly identical directional estimates, in the impossible maze following the local place-to-place metrics would predict that participants point to different locations depending on whether they are biased in a cw or ccw direction. More precisely, as the impossible maze is a disjointed and widened up version of the possible maze direction estimates should show a strong outwards bias, more to the left when following the cw target sequence, more to the right when following a ccw target sequence. Because of the symmetry of the underlying heptagon and decagon for the possible and impossible maze the overall pattern remains the same independent of the current position a participant is performing the task from. Thus, the performance patterns of all sets of trials accomplished by a participant can be merged and averaged (Figure 2, panel C). A more detailed description of the predictions can be found in the results section.

Never had the participants to point the full circle within a set of trials. We decided to restrict the relative corridor distance to a maximum of four as we wanted to prevent participants to notice their own potential outwards biases. For example, imagine participants perfectly recall the local place-to-place metrics along the predefined clockwise target sequence. The global mismatch and the mismatch with the place-to-place metrics of the counterclockwise target sequence should become more and more apparent the closer one gets to closing the circle, thereby coming closer and closer again to one's current location. At some point it becomes easy to compare the estimated location of the counterclockwise, direct neighbor target and the estimated location when following the clockwise place to place metrics. Potentially participants might start to adjust for these mismatches in hindsight during recall, preventing us from drawing conclusions about their stored representation.

In sum our experiment followed a 2 *maze type* (possible vs. impossible) x 2 *order of target sequence* (cw vs. ccw) x 4 *relative corridor distance* (1-4) design with maze type varied across participants and order of target sequence and relative corridor distance varied within-subject. From all seven places within the environment each participant faced four target places following a target sequence along the relative corridor distance one to four (first the direct neighbor then the following three subsequent neighboring corridors), and this was either done in a clockwise or counterclockwise facing order, summing up to 56 individual facing trials. Those were repeated twice across two blocks, adding up to 112 trials per participant in total. Sets of trials were presented randomly within a block. Short pauses of self-determined length were provided after accomplishing seven sets of trials (i.e., 28 pointing trials). Time for self-localization as well as latency and error (angular error between the correct direction to the target following the local corridor-to-corridor metrics and the facing direction of the participant) when facing the target objects were recorded.

After the experiment we asked participants to draw a map of the environment and to fill out a questionnaire. In the questionnaire and in the subsequent debriefing we tried to assess whether participants that learned the impossible maze noticed the global mismatch by asking indirect ("Did you notice anything unusual in the environment?") and direct questions ("Sometimes we visually teleported you from one object location to another. Did you notice?). Sometimes participants reported the impression that the order or number of places or the length of the corridors changed during learning. As this was clearly not the case we attributed such comments to the difficult process of learning the complex environment rather than noticing a global inconsistency. Additionally, we marked down comments made by the participants during the learning phase, the testing phase or short interview in the debriefing phase that indicated detection of global inconsistency (e.g., "I should not be back at the book yet"; "Is this environment possible?"; "There are two places I could point to, where should I face?", "I noticed something was off, but I tried to make sense of it."). To validate their answers, we asked the same questions to the possible maze group. If either of these recordings suggested that a participant felt that the environment does not match up he/she was labelled as someone who "noticed" a mismatch. If not, he/she was denoted as "not noticed".

Results

From the sample we excluded 4% of the data points from pointing error and 4.9% of the data points from pointing latency as those deviated more than ±2 *SD* from a participant's mean performance. As mentioned in the sample description we excluded four participants in the impossible maze group and six in the possible maze group based on their poor pointing performance. This procedure was straight-forward in the possible maze group as targets where clearly anchored in Euclidean space and pointing error defined relative to these locations. If participants mean absolute pointing error was not significantly better than chance (90°) we concluded that no approximately correct representation of the maze was formed and decided for exclusion of the participant. For the impossible maze group different outcomes were possible. We calculated participants pointing error relative to the graph prediction, thus, as a deviation from the correct location when following the local place-to-place metrics. Additionally, we calculated the error relative to a potential global embedding, which should involve the distorted representation of local place-to-place metric to form a Euclidean representation of space. Averaging across all places this should lead to a somewhat even spread of the seven places along a circle, therefore, approximating alignment with the predicted directions in the possible maze group (a heptagon; see following paragraph "Predictions"). Tests for chance level performance were done for both errors, relative to local and global predictions, and participants excluded when they were below chance in both cases. Decisions for exclusion based on the two errors coincided perfectly. The four participants showing chance performance relative to local predictions also showed chance performance relative to global predictions and vice versa. Thirteen trials had to be excluded from the analysis as they were close to an error of 180° (facing away from the correct direction) and their sign of the error differed depending on whether they were calculated relative to local or global predictions. Thus, these trials were ambiguous in their interpretation of whether they reflect a leftwards or rightwards bias relative to a target location. On average participants in the possible maze walked 8 laps (SD = 3.20) through the environment during learning (excluding the laps during the learn-check phase), did 2.00 (SD = 1.60) learn-checks before reaching the learning criteria of 100% correctly recalled places, and spent 32.30 min (SD = 20.87) learning the environment. Full exposure time including the time during the learn-check in the deprived maze was on average 43.89 min (SD = 22.81). Correspondingly, in the impossible maze participants walked 6.63 laps (SD = 1.34), did 1.32 (SD = 0.67) learn-checks and spent 24.89 min (SD = 9.39) learning the environment, and had a full exposure including learn-check time of 32.63 min (SD = 12.61) before continuing with the testing phase. Neither the pure learning time nor the full exposure time including the learn-check time was significantly different between groups, $ts < 1.89$, $ps > .070$.

Facing the same direction?

If space is represented in the form of a Euclidean map participants should face the same direction when being queried the same target independent of the order of target sequence (cw vs. ccw) that is biased within a set of trials. In our experimental setup with seven targets in both of our circular mazes facing the two targets in clockwise order with a relative corridor distance three and four are identical to the two targets with relative corridor distance four and three when faced in counterclockwise order (compare to Figure 3). If the place-to-place metrics of the impossible maze are adjusted to fit a Euclidean maze we would expect participants to form a representation which is approaching a symmetric heptagon. Of course, individual corridors and their neighbors might be susceptible to distinct representational deviations and those might differ across participants. However, as we are averaging across all possible location a participant is pointing from such deviations should be centered out. Therefore, we reckon a symmetric heptagon—similar to the predicted pointing pattern for the possible maze—to be a fair baseline for testing global, Euclidean embedding of our impossible space. In contrast, if a graph-like representation is underlying survey estimates leftwards biases are expected for sets of clockwise target order and rightward biases for sets of counterclockwise target order for the impossible maze group. Therefore, in our first analysis we examined how participants facing direction for relative corridor distance three and four deviate from a symmetric heptagon. The predictions are summarized in Figure 4, left.

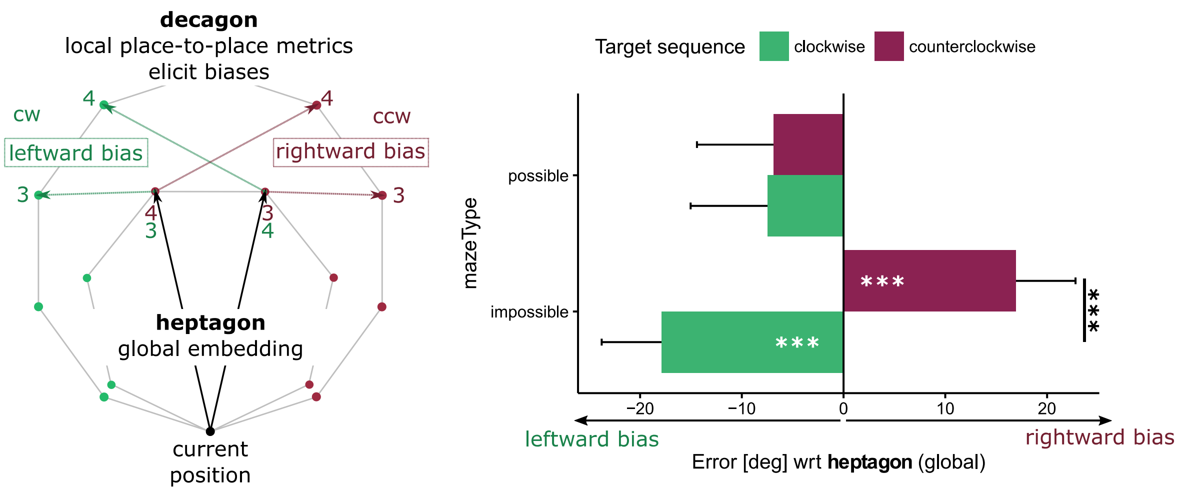


Figure 4. **Left:** Predicted biases when facing the same targets (targets three and four corridors away) when facing clockwise or counterclockwise. In the case of global embedding participants in the impossible maze group should face the same direction independent of biased direction (cw or ccw), thus approximating facing directions that are expected for the possible maze (heptagon). When relying on a graph like representation of local place-to-place metrics leftward and rightward biases are expected for cw and ccw target sequences respectively. **Right:** Averaged facing error for relative corridor distance three and four with respect to predicted headings for a symmetric heptagon are shown for the possible and impossible maze type groups. While the possible maze group shows the expected pattern of pointing along the heptagon and faces the same direction irrespective of the target sequence, clear leftward and rightward biases are found for the impossible maze group. Participants in this group do not point to the same location when being queried to face the same targets. * $p < .050$; *** $p < .010$

Using signed error with respect to a predicted global embedding along a heptagon we conducted an ANOVA with the factors *maze type* (possible vs. impossible) and *order of target sequence* (cw vs. ccw). We included data points from half of each set of trials, namely, only when participants reached relative corridor distance three and four (which included the same target objects for cw and ccw target sequence). Thereby, we left out the first and second trials of each set when facing the neighboring corridor and the second neighboring corridor. No main effect of *maze type* was found, $F(1, 34) = 1.42$, $p = .241$, np2 = .04, but a significant main effect of *order of target sequence,* $F(1, 34) = 6.84$, $p = .013$, np2 = .17, as well as an interaction of *maze type x target sequence* $F(1, 34) = 6.37$, $p = .016$, np2 = .16. When further examining the error pattern for clockwise and counterclockwise sets of trials (visualized in Figure 4, right) post-hoc t-tests revealed that the possible maze group showed similar performance for cw and ccw trials, $t(34) = -0.06$, $p = .949$. However, the error pattern for the impossible maze group differed significantly, $t(34) = -3.63$, $p < .001$. While for trials biased with the clockwise order of target sequence participants yielded a significant leftward bias, different from zero with $t(17) = -2.93$, $p = .009$, trials biased with the counterclockwise order participants showed a significant rightward bias, $t(17) = 4.03$, $p < .001$ (two sided t-tests against the global embedding prediction). Thus, participants in the impossible maze group did not face the same direction in the cw and ccw sets of trials although being queried to face the same targets. Although not shown in the Figure, this pattern was identical for both targets when examined separately. This means, only examining error when facing relative corridor distance three cw in contrast to relative corridor distance four ccw or when facing relative corridor distance four cw in contrast to relative corridor distance three ccw revealed that participants did not face the same direction, ps < .010, but instead exhibited leftward and rightward biases accordingly.

We evaluated whether participants noticed that the environment they learned was impossible (see Procedure). Based on this, eight participants in the impossible maze group were classified as "noticer". The remaining ten participants did "not notice" anything and were surprised when being undeceived post-hoc, stating that everything felt totally normal during the experiment. An ANOVA with the between subject factor *mismatch noticed* (noticed vs. not noticed) and the within-subject factor *order of target sequence* (cw vs. ccw) based on the data of the impossible maze group revealed neither a main effect of nor an interaction with *mismatch noticed*, p's > .103. Only the main effect of *target sequence* was significant, $F(1, 16) = 16.76$, $p = .001$, np2 = .51. Both the participants who "noticed", $t(16) = -3.11$, $p = .007$, and the ones who did "not notice" a mismatch, $t(16) = -2.72$, $p = .015$, differed in the displayed error. Thus, both groups of participants faced different directions albeit being queried the same targets and either showed a leftwards or rightwards bias depending on whether they were biased with a cw or ccw target sequence. Interestingly, also in the possible maze group four out of 18 participants reported to have felt a mismatch as well (e.g., "I felt I was moved to a different position each lap.", "I couldn't make a circle in my mind.") albeit learning a Euclidean space.

Survey estimates based on local metrics?

In the second part of the analysis we aimed to ascertain where exactly participants in the impossible maze point to, more precisely, whether their error pattern fits to the pattern predicted by the local place-to-place metrics experienced during learning. Figure 3, panel B, visualizes the predictions based on the target layout of the possible and impossible maze. Now cw and ccw order of target sequences are normalized and mapped onto each other to represent the expected outward bias instead of the equivalent left- and rightward bias[16]. Taking the heptagon layout of the possible maze as baseline, precise predictions for the expected outward distortions when following the local place-to-place metrics of the decagon can be made for the impossible maze group. For relative corridor distance one an outward bias of ca. 7.71° can be expected and the expected distortion increases by ca. 7.71° across increasing corridor distances up to ca. 30.86° outward bias expected for relative corridor distance four. The two patterns predicted for global embedding approximating a heptagon and graph knowledge following the decagon structure are visualized in the result graph in Figure 5, panel B, with a solid line with a slope of zero and a dashed line with a slope of 7.71° intercepting at zero (representing one's current position), respectively. Normalized error relative to the heptagon as a function of relative corridor distance is depicted, positive values representing outward biases.

First, using the normalized error an ANOVA with the factors *maze type* and *relative corridor distance* showed that the error patterns of possible and impossible maze group differ, supporting the first part of the analysis. The ANOVA yielded a significant main effect of *maze type* $F(1, 34) = 8.07$, $p = .008$, np2 = .19, further supported by post-hoc t-tests showing that the difference of

---

[16] We tested whether the error pattern shown by the two *groups* of participants for inward-outward biases differ between cw and ccw sets of trials by running and ANOVA with the factors *relative corridor distance* and *order of target sequence* for each maze type group separately. For the impossible maze group, no main effect of *target sequence* and no significant interaction with *relative corridor distance* was found, *p*'s > .290, indicating comparable error patterns both for cw and ccw sets of trials. For the possible maze group a significant interaction of *relative corridor distance x target sequence* was found, $F(3, 51) = 3.27$, $p = .029$, np2 = .16. As already implied in Figure 4, right panel, there seems to be a slight, overall leftward bias for the possible maze group, which is reflected in a slight outward bias for cw and an inward bias for ccw sets of trials in the normalized case, leading to the observed interaction. Nevertheless, we decided to merge data for cw and ccw sets of trials. Firstly, the possible maze group is merely considered as a baseline condition and the leftward bias is not strong, secondly, normalization and merging of data is valid for the crucial condition of the impossible maze group.

error patterns is immanent in relative corridor distance two, $t(71.16) = 2.25$, $p = .027$, three, $t(71.16) = 3.16$, $p = .002$, and four, $t(71.16) = 2.61$, $p = .011$, but not yet in distance one, $t(71.16) = 1.23$, $p = .223$. This might be related to the small difference in predicted directions for graph representation vs. Euclidean map of only 7.71° for this corridor distance. Furthermore, a significant main effect of *relative corridor distance* was found, $F(1.89, 64.15) = 4.38$, $p = .018$, np2 = .11, but no interaction with *maze type*, $F(1.89, 64.15) = 1.48$, $p = .237$, np2 = .04 (p-values and degrees of freedom for the following effects were Greenhouse-Geisser-adjusted as Mauchly $p \leq .05$): *relative corridor distance*; *maze type x relative corridor distance*). Taking a closer look at the effect of corridor distance with post-hoc t-tests revealed that significant differences across distances in the outward bias pattern were only found in the impossible maze group. The outward bias increased significantly from relative corridor distance one to three, $t(102) = -3.53$, $p = .004$, and one to four, $t(102) = -3.37$, $p = .006$ (remaining ts ≤ 2.17, ps $\geq .136$). No difference in error pattern across distances could be found in the possible maze group, ts ≤ 1.30, ps $\geq .565$ (Tukey method correction for comparing a family of four estimates).

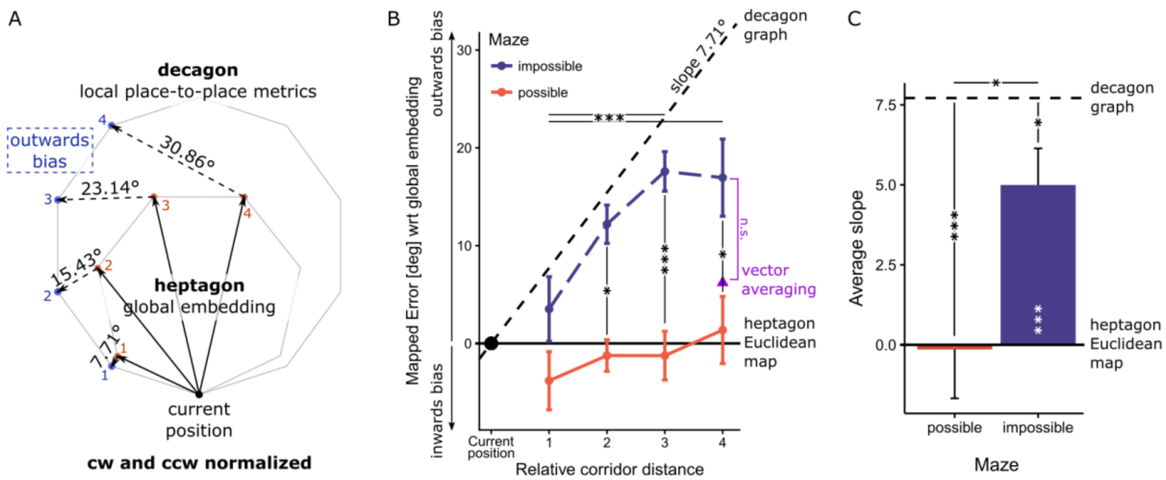


Figure 5. Examination of outward biases. **A|** Predicted outward biases relative to a global embedding when instead following the local place-to-place metrics. Matching up the heptagon and decagon underlying the target layouts for possible and impossible maze a linear increase of the outward bias by ca. 7.71° per additional relative corridor distance can be expected if pointing is based on a local graph representation. **B|** Normalized error (positive values representing an outward bias) across relative corridor distances and the predicted patterns for either the Euclidean map (estimates approaching a heptagon) in black solid line (zero line) or the local graph representation (estimates following corners of decagon) in black dashed line with a slope of 7.71°. **C|** Average slopes extracted from linear regressions of individual performances across relative corridor distances. Average slope of the impossible maze group leans towards the predicted slope for a graph representation but remains significantly different from it.

We again controlled whether there is a difference in the error pattern for participants that “noticed” a mismatch and participants that did “not notice” anything in the impossible maze group. We ran an ANOVA with the factor *relative corridor distance* and *mismatch noticed*, but neither found a main effect, $F(1, 16) = 3.38$, $p = .085$, np2 = .17, nor an interaction with *mismatch noticed*, $F(1.81, 29.01) = 1.45$, $p = .250$, np2 = .08, only a main effect of *corridor distance*, $F(1.81, 29.01) = 5.04$, $p = .016$, np2 = .24. Therefor the patterns shown in both groups are comparable.

In a second step we examined whether the increase across relative target distance found in the impossible maze group is in accordance with the predicted slope when perfectly following local place-to-place metrics. For the full data set of each participant we ran a linear regression of normalized error over the four relative corridor distances. The intercept (distance zero) was hardcoded to zero, as this corresponds to participants’ current position and no deviation would be expected if participants would have to directly point to their own current location. From each linear regression we extracted participants’ individual slope, reflecting the increase in outward bias over the relative corridor distances. A similar procedure was pursued for the possible maze group. Figure 5, panel C, shows the average slopes of both maze groups. The mean slopes differ significantly between groups, $t(31) = 2.69$, $p = .011$, $d = .90$. The average slope of 5° per additional corridor distance for the impossible maze group is significantly higher than a zero slope, which would be expected for global embedding approximating a heptagon, $t(17) = 4.37$, $p \leq .001$. Yet again it is also smaller than the slope of 7.71° expected for a graph representation, $t(17) = -2.38$, $p = .029$. The average slope for the possible maze group of -0.15° is not significantly different from the expected slope of zero for this group, $t(17) = -0.10$, $p = .925$, reflecting a good fit to the heptagon prediction. As expected this slope is significantly smaller than 7.71°, $t(17) = -5.14$, $p \leq .001$.

To examine which prediction fits the data of our impossible maze group best we consulted the Bayesian Information Criterion (BIC) for the fit of two different linear regression models: (1) the local graph model describing the linear relationship across distances with a slope of 7.71° and (2) the Euclidean map model describing the relationship across distances with a slope of 0°. This time we regressed over the averaged data for each participant (thus, the data that underlies Figure 5, panel B, and the previous ANOVA). Again, for both models the intercept at distance zero (current position of participant during pointing) was hardcoded to zero. For calculating the likelihood and based thereon the BIC for the two models, sigma was specified based on the standard deviation of normalized error from the control group, the possible maze group (SD = 19). The local graph model yielded a BIC_{LG} of 639.1, the Euclidean map model a BIC_{EM} of 647.0, resulting in a difference of $BIC_{DIFF} = BIC_{EM} - BIC_{LG}$ = 7.9. Based on the likelihood ratio of the local graph and Euclidean map models we further calculated a general Bayes Factor of BF_{GM} = 52.9. The size of both the BIC_{DIFF} as well as the BF_{GM} correspond to strong evidence that the local graph model fits the data better (according to Raftery, 1995).

What is happening at relative corridor distance four?

As can be seen in Figure 5, panel B, the error pattern of the impossible maze group does not fit perfectly to the local graph model. Most prominent is a flattening of the previously increasing outward bias over corridor distances when switching from corridor distance three to four. In contrast to the other distances (t-tests of error pattern at relative corridor distance 1, 2, 3 against labelled graph prediction, ts < 1.53 , ps > .144) the pattern of distance four is significantly different from the local graph prediction, $t(17) = 2.58$, $p = .019$. However, it is also clearly different from the global embedding prediction, $t(17)$ = 3.15, $p = .006$. This is opening the question as to what is happening at relative corridor distance four and whether there might be strategic or procedural changes in recalling spatial information. Relative corridor distance four is special in a sense. While being biased in a predefined target sequence following cw or ccw order reaching the fourth target is closer for the opposite, non-biased direction in terms of numbers of local entities that must be activated along the path to the target. For example, facing the fourth target in a row along a cw order corresponds to the same object which is only three corridors away from one's current position if following the ccw corridor order. Thus, if following nodes from a graph representation only three instead of four local entities must be activated if taking the opposite "mental route" towards the target.

One reason for the dip could be that individual participants vary in the way they recall the fourth target within a set of trials. For example, sometimes they might be taking the predefined, biased "mental route" while recalling the place-to-place metrics, sometimes they might switch to recalling the fourth target by successively activating local entities along the opposite direction. If this is the case for the fourth target participants should show higher variability in their error pattern for corridor distance four and the time needed for facing should increase compared to the relative corridor distances one to three and compared to the possible maze group. First, we examined the standard deviation of participants individual error pattern across distances (Figure 6, panel A). An ANOVA revealed only a main effect of *relative corridor distance*,

$F(1.91, 64.89) = 10.51$, $p < .001$, np2 = .24, but no main effect of, $F(1, 34) = 0.00$, $p = .994$, np2 < .01, or interaction with *maze type*, $F(1.91, 64.89) = 0.09$, $p = .904$, np2 < .01. In both groups of participants noise in the data increased with increasing corridor distance. Post-hoc t-tests yielded significant differences in standard deviation averaged across both groups between corridor distance one and three, $t(102) = -3.62$, $p = .003$, one and four, $t(102) = -5.15$, $p \leq .001$, and two and four, $t(102) = -3.77$, $p = .002$, remaining $ts \leq 2.23$, $ps \geq .120$. No particular increase of noise for distance four that goes beyond the linear increase that is visible throughout the whole range of distances can be detected. Also, the averaged individual standard deviation did not differ between groups, for neither of the four relative corridor distances, also not for distance four, $ts \leq 0.14$, $ps \geq .888$. The constant increase can be well explained by the successive activation of and associated error accumulation across local memory units along the way, which would be predicted by graph representations. Second, we ran a similar ANOVA but on pointing latency (Figure 6, panel B). Again, only a main effect of *relative corridor distance* was found, $F(1.43, 48.59) = 23.68$, $p < .001$, np2 = .41, but no main effect of or interaction with maze type, $Fs \leq 0.66$, $ps \geq .422$. For both maze types facing the first target in a set of trials took significantly longer than facing to the succeeding targets, $ts \geq 6.11$, $ps \leq .001$. The remaining post-hoc tests between relative corridor distance two to four (including latency difference between relative corridor distance three and four) were not significant, $ts \leq 1.09$, $ps \geq .698$.

In sum, both approaches suggest that participants do not change their procedure to estimate survey relations when facing the fourth target compared to the previous targets within a set. Besides error accumulation across local entities there is no particular additional increase in noise for distance four, instead participants seem to be rather constant in their behavior and their decision for a target to face to. This speaks against the potential strategy that a participant alternates between the biased sometimes the unbiased direction (cw or ccw).

Besides intra-individual differences in survey estimates also between-subject variation in responding to the fourth target can lead to the examined flattening of the error pattern over relative corridor distance. Figure 6, panel C, shows a more detailed version of Figure 5, panel B, where the distribution of the average performance of each participant across all distances is depicted as well. For the impossible maze group the range of responses is highest for relative corridor distance four. The standard deviation of the averaged values across distances are: $SD_1 = 13.56$, $SD_2 = 10.93$, $SD_3 = 15.25$, $SD_4 = 25.37$. While a few participants show a very strong outward bias one participant even shows an inward bias. Hartigans' dip test for unimodality/multimodality, however, did not yield indications for bimodality or multimodality though, not for the distribution of error pattern for distance four nor the other three distances, *ds* ≤ 0.08, *ps* ≥ .376. Thus, it does not seem appropriate to divide the sample in groups of different responders, for example one group that points along the local graph prediction, some that show adjustments towards global embedding.

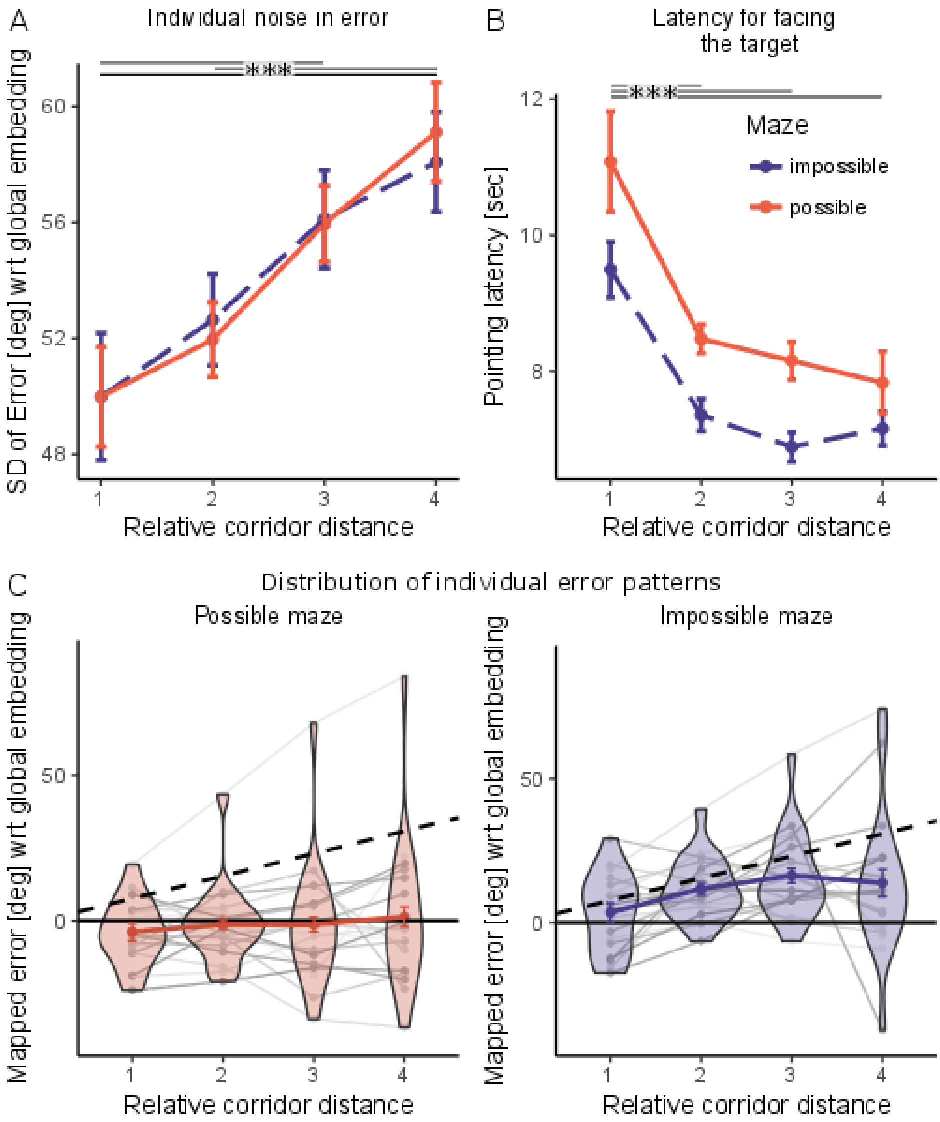


Figure 6. Examining what is happening at distance 4. **A|** Dispersion of each participant's responses about their own mean direction estimate. No particular increase for relative corridor distance four that goes beyond the linear increase across all distances is shown. **B|** Time needed to estimate target directions. Relative corridor distance four is not particularly more time consuming. **C|** Detailed version of Figure 5B including (in shades of grey) individuals average performance across distances and the distribution of participants average pointing patterns for each corridor distance in the form of a violin plot. In the impossible maze group, the highest standard deviation of inter-individual error variability can be found for relative corridor distance four.

Although the analysis of latency and individual standard deviation does not indicate that there is a general change in the way a participant estimates survey relations for the fourth target in the impossible maze group, we also considered possibilities alternative to the biased local graph or global embedding models. One possibility stated above already is that participants switch to estimate the targets' location along the opposite direction of the mental route (e.g., being biased in cw target sequence but recalling target four along the ccw place-to-place metrics). If this would be the case a strong inward bias would be expected, which is clearly not reflected in the data. Alternatively, albeit being biased along one direction two vectors to the same target might have been estimated along both the cw and ccw direction and then averaged for the fourth target (i.e., averaging vectors to the biased cw4 location and the additionally activated ccw3 or for the other direction averaging vectors to the biased ccw4 location and the additionally activated cw3). In this case a weak outward bias of 6.21° would be predicted for distance four, which is visualized as a violet triangle in Figure 5, panel B. The error pattern of the impossible maze group at relative corridor distance four is not significantly different from this prediction, $t(17) = 1.27$, $p = .223$. Yet, running a Bayesian analysis on the difference of the predicted and the shown error pattern only revealed a weak support for the Null-hypothesis ($BF_{01} = 2.07$, anecdotal evidence according to Wagenmakers et al., 2011). Note that vector averaging of the previous corridor distances 1_{biased} | $6_{unbiasedOpposite}$, 2_{biased} | $5_{unbiasedOpposite}$, 3_{biased} | $4_{unbiasedOpposite}$ would have led to inward biases.

It should be noted that when running the slope analysis (linear regression on participants individual data points) again while excluding relative corridor distance four the average slope increases to 5.75° and is not anymore significantly different from the 7.71° slope predicted for local graph representations, $t(17) = -1.65$, $p = .118$. But running a Bayes analysis reveals only anecdotal evidence for the Null-hypothesis, $BF_{01} = 1.33$, namely, that there is no difference between the calculated slope and the expected slope for the labelled graph model.

General Discussion

In our study we set out to examine which memory structure underlies survey estimates within a complex environment, more precisely, we compared two competing approaches that describe how spatial survey knowledge is represented. The Euclidean map approach implies that locations we experience in space are assigned to discrete points in a mental coordinate system, thus, in a globally consistent format. In comparison, labelled graph approaches suggest that we store units of local information only, meaning, we memorize distance and angle between adjacent places. Those individual place-to-place metrics can be globally inconsistent. Variations of how the individual pieces are recalled can results in different estimates of direction and distance to a target. We contrasted both approaches by having participants learn either of two circular environments, a possible Euclidean space or an impossible non-Euclidean space, where local place-to-place metrics to not match up on the global scale. The possible maze group was well able to learn the complex seven-corridor environment and pointed consistently to targets along the underlying heptagon distribution of the places irrespective of the biased direction. In contrast, the impossible maze group did not face the same direction to the same target. Depending on whether they pointed towards the targets in a cw or a ccw order along the circular layout, leftward and rightward biases were shown respectively. Condensed these distortions reflect an outward bias in the direction of the local place-to-place metrics of the impossible maze. The pattern is better explained by the labelled graph model than the global embedding model since the latter would have required the mental representation to be brought into a coherent format, hence, making the impossible environment possible in memory in Euclidean terms. In other words, the Euclidean map format should have distorted the local place-to-place metrics navigators experienced in a way to match up globally, leading to uniform pointing patterns.

Our first analysis that concentrated on the two targets distant three and four corridors away showed a mean leftward bias of 17.89° when being biased in cw sequence and mean rightward bias of 16.92° when being biased in ccw sequence relative to an even spread around a heptagon which was predicted for global integration. For these trials from each of the seven locations within the environment the exact same target identities had to be estimated with the only difference that participants were pointing to the targets in a different order. Clearly, participants in the impossible maze group did not face the same direction despite being asked the same targets while standing at the same spot. This pointing pattern conflicts with the metric postulate of positivity, which implies that each unique place can only be assigned to a single location in the mental map. The positivity further specifies that the distance from any point to itself must be zero and no other place can be assigned the same coordinates (e.g., Beals et al., 1968; McNamara & Diwadkar, 1997; Warren et al., 2017). The divergence of pointing directions for the impossible maze group, however, indicates that there is more than one true location for the targets pointed to.

Can the category-adjustment model of spatial coding (Huttenlocher et al., 1991; N Newcombe et al., 1999) or the contextual-scaling model (McNamara & Diwadkar, 1997) explain or results just as they are able to explain asymmetries previously found in distance estimation between pairs of places (e.g., Burroughs & Sadalla, 1979; McNamara & Diwadkar, 1997; Sadalla et al., 1980) thereby refuting violations of Euclidean embedding? We reckon they can't. Let's consider an example using the category adjustment model. During learning categories or regions should be formed in addition to a Euclidean representation of space. This involves merging of a subset of places into clusters. Typically, this clustering is assumed to follow salient information from the surrounding (the typical example is that vertical and horizontal imagined axes divide a computer screen or a map into four quadrants, each of which is then possessing a category prototype). In our case it is hard to make claims about which places are clustered, but for now let's assume lamp, book and key form cluster 1 (the starting area), duck and comb cluster 2 (connected by sharp turning angle) and dice and shoe cluster 3 (connected by obtuse turning angle). If a participant must estimate the direction to the key (target) while standing at the shoe (reference) (see also Figure 3) the bias model holds that the category prototype of the target (key belongs to cluster 1) should bias its location estimate. This bias now differs when pointing the other way around, hence, from the key (reference) to the shoe (target) as now a different category prototype (shoe belongs to cluster 3) is activated and integrated in a Bayesian manner with the location information from the Euclidean map. Hence, asymmetries occur. Similarly, also in the contextual scaling model the asymmetry arises from interchanging reference object and target object. Hence, both approaches are able to explain specific asymmetries found for distance estimations (e.g., Burroughs & Sadalla, 1979; McNamara & Diwadkar, 1997; Sadalla et al., 1980) and route selection (Stern & Leiser, 2010) by the activation of different contexts or prototypes specific to distinct places. Our study, however, does not rely on interchanging reference and target objects, but instead utilizes sets of trials that have the same reference object (i.e., current position of participant) and the same target objects (relative corridor distance three and four) and only varies what objects are faced to beforehand (cw vs. ccw). Two arguments can be made against the explanatory power of the two models with regards to our study. First, as we average across all sets of trials for the analysis (i.e., merging sets of trials where pointing is done from the book to the dice and from the dice to the book) all potential prototype and category effect should average out. Still, we observe bimodal estimates, hence, a violation of the positivity postulate. Second, the bimodality does not seem to be an artifact of this averaging process. When considering only subsets of trials with the same reference object, hence, when only looking at the pointing patterns made from the book, or those made from the key, or from the dice etc. to the targets three/four corridors away, the bimodality pattern of leftward and rightward bias for clockwise and counterclockwise target sequences that was found in our first analysis was reflected all subsets. In four out of seven reference objects (i.e., current location) this difference in estimated target direction was statistically significant and for the remaining three out of seven the descriptive values went into the predicted direction (additional analysis not reported in the results). Thus, despite constant reference-target pairings, or in other words, despite the same context or category prototype that is activated, the direction estimates still differed.

One could argue that the context-scaling model could be adjusted by adding the assumption that previous estimates of targets at relative corridor distance one and two are considered for subsequent estimates as well. Each estimated target possesses its own saliency (note that differences in saliencies between places should be averaged out as all sets of trials are merged). Each previous target might bias the subsequent estimates towards their location. We do not think this is a likely explanation for our results, as in this case we should see similar outward biases in the possible maze group as well. In sum, we reckon current bias models to be insufficient to explain the results of our study. Our results line up nicely with asymmetries found in direction estimates that are facilitated when done route forward compared to route backward but are independent of the actual reference and target objects (e.g., Meilinger, Henson, Rebane, Bülthoff, & Mallot, 2018; Meilinger, Strickrodt, et al., 2018). We consider these route direction effects in survey estimates to be just as difficult to be accounted for by bias models as our results. Correspondingly, our results suggest the use of place-to-place metrics of a labelled graph representation, and that we succeeded in determining the use of only a subset of such memory units for each set of trials with our manipulation of target sequence.

In our second analysis we ascertained that the outward bias (relative to facing along a symmetric heptagon) that was observed in the impossible maze group was well captured by the impossible local place-to-place metrics experienced during learning. Despite no perfect match was achieved the outward bias increased steadily as predicted by the labelled graph prediction up until relative corridor distance three (the flattening at distance four will be discussed further below). This captures how precise (on average) local metrics are perceived and stored in memory. Our results provide statistical support for previous studies that indicated violation of metric postulates using impossible mazes. For example, findings by Kluss and colleagues (2015) indicated that that the angular metrics experienced in an impossible triangle environment were represented as they were experienced and not embedded globally. However, the authors did not succeed in substantiating this with inference statistical analyses (i.e., no significant difference from 180° sum of inner angles). Indeed, a similar problem occurred in a pre-study to the one presented here. It involved learning either of two circular six-corridor mazes, a possible and an impossible version of it. The corridors or the mazes were arranged in a much simpler and regular form than the ones used in this study (i.e., a slightly distorted regular hexagon) and the differences in outward bias predicted for the impossible place-to-place metrics compared to the possible hexagon where too small. In the end, despite compelling biases in the descriptive values of the pointing patterns the predicted differences were too small to be detectable in the variable performance of the participants. Hence, future research on this should be cautious to choose environments and experimental setups that indeed allow detection of differences within the typically very noisy pointing data. Coming back to previous impossible maze studies, our study further implies that the rips and folds and pointing biases towards wormhole locations in the study of Warren and colleagues (2017) are likely not based on a distorted yet Euclidean representation of space. Instead and taken together a labelled graph that specifies metrics on the local place-to-place level is the most feasible representational structure to explain the current as well as previous studies.

Previous studies have shown that the number of corridors residing between one's current location and the target is correlated with the time needed for making the estimate (e.g., Meilinger et al., 2016; Meilinger, Strickrodt, et al., 2018; Strickrodt et al., 2018). Such corridor distance effects (which cannot be accounted for by straight-line Euclidean distance effects between locations) were typically interpreted as the successive processing of interlinked places following the mental route along a graph representation from one's current location towards the target. Each additional memory unit (i.e., a corridor) along the mental route must be activated first, thus, results in latency increase for making the estimate. Our study both uses and supports this notion. First and foremost, we used this successive activation to trigger which direction along the mental graph of our circular environment is used to come up with an estimate (cw or ccw). By predefining the sequence of targets within a set of related trials we successfully prescribe which of the stored metrics are used for the survey estimate leading to observed outward biases that nicely follow the predicted local metrics.

Second, from previous studies we know that when pointing to a number of related, neighboring targets subsequent pointings seem to be based on previous pointings (Meilinger, Strickrodt, et al., 2018). More precisely, it seems that when pointing from neighbor to neighbor only the difference between the new target and the previous target is calculated instead of discarding all retrieved information from previous targets and re-starting the entire recalling process again. This is reflected in our results as well, importantly, both for the possible as well as the impossible maze group. Albeit increasing relative corridor distance across a set of trials pointing latency remained stable after the first estimate[17]. This implies that for every new target only one more memory unit was activated and integrated with the previous estimate to come up with the next one. Hence, subsequent estimates are not independent. Only when being teleported to a new location at the beginning of a set of trials led to reiteration of the entire process (i.e., again all the way from my current location to the target) from scratch. Note that the longest response latencies were found for the first trial within a set. Here, participants might partly still be in their final phase of self-localization when trying to relate one's current position and corridor orientation to neighboring corridors. Additionally, the first trial is the one with largest uncertainty of which target is going to be asked—either the cw or the ccw neighbor corridor. As soon as this first estimate is done all subsequent targets can be predicted unambiguously (always the next neighbor along the predefined order) and the relation between one's current position and the first target serve as a good reference for one's overall orientation within the environment, making subsequent estimates faster.

A third argument for successive activation of local memory units along the graph can be made based on participants intra-individual standard deviation of pointing error. Individuals standard deviations increased linearly across the sequence of trials in a set both in the possible and the impossible group. Such a pattern can be well explained by error accumulation during recall as each new memory unit that is accessed and incorporated into previous estimates adds noise to the process. In a Euclidean map estimates should be independent of each other. Each direction to a new target can be estimated

[17] Note that this procedure contrasts with studies showing corridor distance effects on latency. Such distance effects seem to be only found when there is no prior knowledge activated in working memory where one's current position is brought into direct reference to another target, thus, corridor distance effects on latency seem to be shown only directly after teleportation to a new location.

again from scratch with similar ease making it unnecessary to rely on previous estimates. Hence, no noise accumulation should occur.

What is happening at relative corridor distance four? Indeed, the outward bias produced by distance four seems to flatten out compared to what would be expected looking at relative corridor distance one to three, suggesting a switch in retrieval procedure. If the opposite, non-biased target sequence was suddenly used for this estimate we would have expected a strong inward bias, which we clearly did not observe. Alternatively, participants might average both vector estimates based on the cw and ccw graph sequence. While such a behavior would predict an inward bias for relative corridor distance one to three, indeed a slight outward bias would be predicted for relative corridor distance four. We only found anecdotal evidence for a fit to this predicted bias for the fourth corridor distance. However, if we hypothesize a switch in the retrieval procedure for relative corridor distance four, for example by suddenly switching to vector averaging, this should lead to additional estimation processes along the non-biased corridor sequence, none of which was activated and maintained in working memory before within the set of trials. This should lead to an increase in latency for relative corridor distance four compared to the previous one. Unfortunately, we don't find this further support in our data.

An alternative explanation that could partly explain the flattening for distance four was, that individual participants dealt with the two conflicting vector estimates from cw and ccw target order in a variable fashion, meaning, they sometimes (maybe more often) chose the biased estimate leading to an outward bias sometimes the unbiased estimate leading to an inward bias. This should yield higher variability in the participants response pattern for distance four. However, no clear indication for such a behavior could be found. Individuals standard deviations increased linearly in both groups across all distances without a particular increase from relative corridor distance three to four. Distinguishing here whether the high intra-individual standard deviation is due to noise accumulation or due to variations in choosing the cw or ccw estimate for corridor distance four unfortunately will not be possible here. If the flattening pattern at distance four cannot be properly accounted for by intra-individual variations in the retrieval process, maybe inter-individual differences in handling the situation of a steadily closing circle and conflicting information of cw and ccw place-to-place metrics might be driving this change. While some participants might stick to following the local place-to-place metrics, others might switch to a vector averaging or responding based on the opposite non-biased corridor order. Looking at the distribution of in-/outward biases between subjects showed that the highest deviation can be found for relative corridor distance four. There seem to be a few "outliers" with strong outwards and inwards biases, yet it seems not feasible to divide participants into reliable groups of switchers vs. non-switchers, as there is no clear bimodal distribution of responders. Also looking at the individual patterns over distances it seems that the response pattern for relative corridor distance four is not corresponding to similar behavior in the other distances. Our analysis remains to be somewhat inconclusive of what exactly is happening at distance four. Different scenarios are possible (e.g., vector averaging), which however all revert to a spatial representation in the form of a local graph structure and cannot be explained by global embedding into a Euclidean map.

We compared whether our observed patterns for the impossible maze group differed between participants that noticed a global mismatch and participants that did not. Firstly, we seem to have introduced a global mismatch so large that it is noticeable to nearly half of the participants. In the wormhole study by Warren and colleagues (2017) 0-9% of participants noticed the 90° rotation of the environment. Suma and colleagues (2012) found that roughly 14% of participants noticed that symmetric rooms connected by an adjacent, straight corridor overlapped in an impossible manner. Authors further showed that detection rate seem to be dependent on room overlap and room size. Reasons for the perceptibility of our impossible architecture could be the strong rotation of the impossible maze by 108° in every lap and the visual elongation of the connection between corridors. The latter was introduced to provide a better understanding of the angle of connectivity between corridors (i.e., longer corners = sharper angles), but could have led to the feeling of overlapping corridors in the case a participant consciously tried to merge corridor information into a globally consistent format. Especially one corridor connection (between duck and comb) is very long and reaching inside the circular layout and far beyond the underlying decagon midpoint, thereby potentially drawing attention to violation of a possible space. Importantly, it should be noted that while identifying eight *noticer* in the impossible maze group we also found four participants in the possible maze group that noticed something albeit learning a Euclidean space. Thus, part of the effect of noticing a global mismatch in the environment might simply be attributed to the fact that learning took place in a VR setup which is close to but of course not 100% equal to natural space. Additionally, participants also might have higher alertness to being tricked as they partook in a psychological experiment, walking an unusually complex environment under a constraint learning procedure. Albeit the uncertainty of how much of the *noticer* effect can indeed be attributed to the impossible maze itself we found it valid to compare both groups. Interestingly, we never found significant differences between the two groups. *Noticer* did not seem to account for the mismatch by adjusting their pointing estimates to match up globally. We can only speculate why this is the case. Potentially they grasped the idea of the experiment and went along with it. Alternatively, they did not know how else to account for the mismatch and decided simply to use what knowledge is available for them. Either way it remains that participants seem to base their estimates on local metrics, whether noticing impossibility or not.

As a potential mechanism for enabling global metric embedding we proposed the recalibration of the path integration system that updates the difference vector to all visited landmarks along the circular environment (compare Wang, 2016). While walking participants could have updated a number of vectors each pointing to one of the objects that had to be learned. Upon first walk through the impossible maze the book is reached again even though the updated vector towards the book may indicate otherwise. Also, the vector to, for example, the key should point elsewhere but the neighboring corridor after the book. These vectors could then have been corrected and adjusted accordingly. Another approach stated by Mallot and Basten (2009) specifies that global embedding could be achieved not necessarily during encoding but by mental triangulation on the local place-to-place metric stored in long-term memory. These are compared, coordinated and adjusted to remove global inconsistencies. Albeit theoretically possible and plausible with the amount of exposure participants have to the seven-corridor environment, revisiting the original book corridor at least six times, our results suggest that neither updating and recalibration nor mental triangulation on the stored metrics took place. Regarding the updating hypothesis, indeed, most studies that showed recalibration of the path integrator used open spaces (e.g., Jayakumar et al., in press; Tcheang et al., 2011) in contrast to the multi-corridor environment used in our study. Further, leaving a local environment was found to impair the updating of objects left behind (Wang & Brockmole, 2003a). Considering mental triangulation, the number of places and stored local metrics might have been too high and mental calculations thereon too complex to be realizable in working memory. We cannot preclude that global embedding was not attempted based on either of these mechanisms. If attempted, it wasn't successful in yielding a globally embedded Euclidean map representation.

Taken together, our study implies that the processing of survey estimates is based on a graph structure and is bound to a successive re-activation of the interlinked places following a mental route along the graph from one's current location towards the target. Each estimate is generally transient but can be used at least for a short while to base subsequent estimates on it. Importantly, the metrics stored in this graph are probably purely local and do not need to be adjusted to match globally.

Accounting for contradicting arguments

We are aware that the merging of the impossible local metrics to fit a symmetric heptagon is a rather strong claim for testing global embedding. Indeed, if places are brought into globally consistent relations we do not object the possibility that unique distortions of length and angle of individual corridors or connections may occur and differ across participants and across the neighboring corridors of our environment. However, because of the circular nature of our environment there are clearly limits to such unique distortions. The seven places revisited again and again along the circular environment must at least roughly be spread across a circle or ellipse If global embedding is achieved the object layout should thereby form a globally consistent map of a non-symmetric heptagon, rather than, for example, a somewhat straight line without a viable connection between the first and last corridor. As we averaged across all current positions and targets such unique distortions of non-symmetric heptagon layouts should average out to a somewhat even spread across a symmetric heptagon. Thus, if participants had formed a Euclidean representation but with strong distortions we would have expected participants error patterns to vary widely around the zero line of Figure 4 and 5 but on average not deviate significantly from it. Therefore, we consider the symmetric heptagon which is also underlying the place layout of the possible maze to be a fair and valid prediction to test for potential global embedding in the impossible maze group.

By having participants learn a non-Euclidean, impossible space we violated the coherence of spatial cues naturally encountered in the real world. How can we be sure that we did not alter the natural, spatial learning processes simply by introducing a global mismatch with our impossible maze manipulation. In other words, if humans usually form Euclidean mental maps of navigable space, how can we argue they do not if we have prevented them from forming one by making them learn a space they would never experience in everyday life? Our results might merely reflect that a different learning procedure was triggered in the impossible maze group. We don't think that this is likely. If fundamentally different processes would take place and the nature of representation differ significantly between both groups, this should reflect in their learning behavior and testing performance. First, learning time and the average laps through the environment during learning was comparable for the possible and impossible maze group. We even had to exclude a few more subjects from the possible maze group because of chance performance in the task. Second, looking at the time needed for pointing and the intra-individual variation in the error pattern showed no significant difference between the two groups. Instead in both groups the variance in direction estimates increased linearly across the relative corridor distances. A Euclidean map could not account for such an increase as locations should be assigned unambiguously to a pair of coordinates. Even if these coordinates are not reflecting the external world correctly, participants should at least always face the same direction irrespective of relative corridor distance, resulting in constant standard deviation. Instead increase in variance can be explained by error accumulation across more and more local units that are accessed and incorporated into the estimation process. Thus, our results speak for the formation of a local graph representation and similar processes underlying direction estimates in both groups.

The data we present here is not based on a long-term navigation study with repeated exposure to the same environment. Our participants learned the space within a single session, experiencing it a minimum of half an hour with clear instructions on the later task. In contrast to everyday life navigation full concentration to all spatial properties of the environment is ensured and only competent navigators were excluded into the analysis. Albeit being able to perform a survey task our results suggest that both groups do not revert to a Euclidean map where each place is already brought into direct reference with the other places. Instead, reference between one's current position and the target seems to be constructed online during testing. Nevertheless, we cannot exclude the possibility that with sufficient exposure to the impossible maze global embedding might take place, equalizing error patterns for both groups. Indeed, there are studies showing that alignment with a global main orientation facilitates retrieval of survey knowledge in navigable space (e.g., Strickrodt et al., 2018; Tlauka et al., 2011; Wilson & Wildbur, 2004; Wilson et al., 2007). Such results suggest the formation of global Euclidean maps that relate more than only a single corridor or room but integrate spatial information from multiple local places into a coherent format. However, besides evidence for global embedding, effects of local memory units that are accessed successively remained in the study of Strickrodt and colleagues (2018) (see also Meilinger et al., 2013). The question arising here is why participants can't make use of this global embedding by simply reading-out coordinates from the Euclidean map?

Why do they still follow the local corridor connectivity experienced during learning instead? It remains a subject of further investigation and theoretic devotion to understand whether evidence for orientation dependent facilitation of survey estimates following a global direction indeed reflect a global embedding into a Euclidean mental map. Indeed, Strickrodt and colleagues (2018) proposed an alternative explanation, namely, the representation of a general reference direction, an additional vector attached to each local memory unit that specifies a constant global direction across multiple local spaces. Despite these ambiguities our results emphasize two things. First, local graph representations seem to be a core medium to solve survey tasks and, second, that a global, Euclidean map is not necessary to conquer navigational challenges like shortcutting or straight-line pointing successfully.

Conclusion

Although we can assume that similar processes do underly survey estimates both in the possible and the impossible maze group—as indicated by similar pointing latency and noise in the data—the direction estimates made by the impossible maze group strongly violate the metric postulate of positivity. Despite being queried to point out the direction to the same target the estimated direction differed and depended on the predefined path we triggered. Clearly, the impossible space that was experienced was not embedded into a global format such as a Euclidean mental map. Instead, the observed pointing pattern could be well described by the predicted outward biases when following a predefined "mental route" along a labelled graph representation. Basing estimates on such a graph structure seems to follow a successive activation of all local memory units along parts of the graph and the local metrics that specify their connectivity. In sum, our results indicate that our mental representation of objects in navigable space consists of a number of piece-wise, interconnected information (Chrastil & Warren, 2014; Meilinger, 2008; Warren et al., 2017), not a large mental reference system which incorporates the metric relations between all places in a Euclidean sense (e.g., Gallistel, 1990; Ishikawa & Montello, 2006; O'Keefe & Nadel, 1978).

Acknowledgments

We would like to thank Joachim Tesch and Adam Hersko-RonaTas for help with the virtual environment setup and Gregory Dachner and Lauren Franklin for the audio recordings for the testing phase.

References

Beals, R., Krantz, D. H., & Tversky, A. (1968). Foundations of multidimensional scaling. *Psychological Review*, *75*(2), 127–142. http://doi.org/10.1037/h0025470

Burroughs, W. J., & Sadalla, E. K. (1979). Asymmetries in Distance Cognition. *Geographical Analysis*, *11*(4), 414–421. http://doi.org/10.1111/j.1538-4632.1979.tb00709.x

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375. http://doi.org/10.1037/0033-295X.114.2.340

Byrne, R. W. (1979). Memory for Urban Geography. *Quarterly Journal of Experimental Psychology*, *31*(1), 147–154. http://doi.org/10.1080/14640747908400714

Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian Integration of Spatial Information. *Psychological Bulletin*, *133*(4), 625–637. http://doi.org/10.1037/0033-2909.133.4.625

Chrastil, E. R., & Warren, W. H. (2013). Active and passive spatial learning in human navigation: acquisition of survey knowledge. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(5), 1520–37. http://doi.org/10.1037/a0032382

Chrastil, E. R., & Warren, W. H. (2014). From Cognitive Maps to Cognitive Graphs. *PLoS ONE*, *9*(11), e112544. http://doi.org/10.1371/journal.pone.0112544

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*. http://doi.org/10.1038/nature08704

Ekstrom, a D., Kahana, M. J., Caplan, J. B., Fields, T. a, Isham, E. a, Newman, E. L., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*(6954), 184–188. http://doi.org/10.1038/nature01955.1.

Farrell, M. J., & Robertson, I. H. (2000). The automatic updating of egocentric spatial relationships and its impairment due to right posterior cortical lesions. *Neuropsychologia*, *38*(5), 585–595. http://doi.org/10.1016/S0028-3932(99)00123-2

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://doi.org/10.1037/0278-7393.31.2.195

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. http://doi.org/10.1038/nature03721

Huttenlocher, J., Hedges, L. V, & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological*

Review, *98*(3), 352–376. Retrieved from http://psycnet.apa.org/journals/rev/98/3/352/

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://doi.org/10.1016/j.cogpsych.2005.08.003

Jacobs, J., Kahana, M. J., Ekstrom, A. D., Mollison, M. V, & Fried, I. (2010). A sense of direction in human entorhinal cortex. *Proceedings of the National Academy of Sciences*, *107*(14), 6487–6492. http://doi.org/10.1073/pnas.0911213107

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X.-X., … Kahana, M. J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience*, *16*(9), 1188–1190. http://doi.org/10.1038/nn.3466

Jayakumar, R. P., Madhav, M. S., Savelli, F., Blair, H. T., Cowan, J., & Knierim, J. J. (n.d.). Recalibration of path integration in hippocampal place cells. http://doi.org/10.1101/319269

Kluss, T., Marsh, W. E., Zetzsche, C., & Schill, K. (2015). Representation of impossible worlds in the cognitive map. *Cognitive Processing*, *16*(S1), 271–276. http://doi.org/10.1007/s10339-015-0705-x

Kosslyn, S. M., Pick, H. L., & Fariello, G. R. (1974). Cognitive maps in children and men. *Child Development*, *45*(3), 707–716. http://doi.org/10.1111/1467-8624.ep12147012

Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. a. (1993). Nonvisual navigation by blind and sighted: Assessment of path integration ability. *Journal of Experimental Psychology: General*, *122*(1), 73–91. http://doi.org/10.1037/0096-3445.122.1.73

Mallot, H. A., & Basten, K. (2009). Embodied spatial cognition: Biological and artificial systems. *Image and Vision Computing*, *27*(11), 1658–1670. http://doi.org/10.1016/j.imavis.2008.09.001

Marchette, S. A., Ryan, J., & Epstein, R. A. (2017). Schematic representations of local environmental space guide goal-directed navigation. *Cognition*, *158*, 68–80. http://doi.org/10.1016/j.cognition.2016.10.005

Martin, J. F., & Thomson, J. A. (1998). Automatic Spatial Updating during Locomotion without Vision. *The Quarterly Journal of Experimental Psychology Section A*, *51*(3), 637–654. http://doi.org/10.1080/713755776

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology*, *18*(1), 87–121. http://doi.org/10.1016/0010-0285(86)90016-2

McNamara, T. P., & Diwadkar, V. A. (1997). Symmetry and asymmetry of human spatial memory. *Cognitive Psychology*, *34*(2), 160–190. http://doi.org/10.1006/cogp.1997.0669

McNamara, T. P., Sluzenski, J., & Rump, B. (2008). Human Spatial Memory and Navigation. In H. L. Roediger III (Ed.), *Learning and Memory: A*

Comprehensive Reference (Vol. 2, pp. 157–178). Oxford: Elsevier. http://doi.org/10.1016/B978-012370509-9.00176-5

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* (Vol. 5248 LNAI, pp. 344–360). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-87601-4_25

Meilinger, T., Henson, A., Rebane, J., Bülthoff, H. H., & Mallot, H. A. (2018). Humans Construct Survey Estimates on the Fly from a Compartmentalised Representation of the Navigated Environment. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI. 11th International Conference, Spatial Cognition 2018. Lecture Notes in Artificial Intelligence*. Springer.

Meilinger, T., Riecke, B. E., & Bülthoff, H. H. (2013). Local and Global Reference Frames for Environmental Spaces. *Quarterly Journal of Experimental Psychology*, *67*(3), 542–569. http://doi.org/10.1080/17470218.2013.821145

Meilinger, T., Strickrodt, M., & Bülthoff, H. H. (2016). Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation. *Cognition*, *155*. http://doi.org/10.1016/j.cognition.2016.06.003

Meilinger, T., Strickrodt, M., & Bülthoff, H. H. (2018). Spatial Survey Estimation Is Incremental and Relies on Directed Memory Structures. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI. 11th International Conference, Spatial Cognition 2018. Lecture Notes in Artificial Intelligence* (Vol. 11034). Springer.

Moar, I., & Bower, G. H. (1983). Inconsistency in spatial knowledge. *Memory & Cognition*, *11*(2), 107–113. http://doi.org/10.3758/BF03213464

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(1), 142–157. http://doi.org/10.1037/0278-7393.30.1.142

Müller, M., & Wehner, R. (1994). The hidden spiral: systematic search and path integration in desert ants, Cataglyphis fortis. *Journal of Comparative Physiology A*, *175*(5), 525–530. http://doi.org/10.1007/BF00199474

Muryy, A., & Glennerster, A. (2018). Pointing Errors in Non-Metric Virtual Environments. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI. 11th International Conference, Spatial Cognition 2018. Lecture Notes in Artificial Intelligence*. http://doi.org/10.1101/273532

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Report Development of Cue Integration in Human Navigation. *Current Biology*, *18*, 689–693. http://doi.org/10.1016/j.cub.2008.04.021

Newcombe, N., Huttenlocher, J., Sandberg, E., Lie, E., & Johnson, S. (1999). What do misestimations and asymmetries in spatial judgement indicate about spatial representation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 986–996. http://doi.org/10.1037/0278-7393.25.4.986

Newcombe, N., & Liben, L. S. (1982). Barrier effects in the cognitive maps of children and adults. *Journal of Experimental Child Psychology*, *34*(1), 46–58. http://doi.org/10.1016/0022-0965(82)90030-3

Nilsson, N. C., Peck, T., Bruder, G., Hodgson, E., Serafin, S., Whitton, M., … Rosenberg, E. S. (2018). 15 Years of Research on Redirected Walking in Immersive Virtual Environments. *IEEE Computer Graphics and Applications*, *38*(2), 44–56. http://doi.org/10.1109/MCG.2018.111125628

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.

Ruddle, R. A., Howes, A., Payne, S. J., & Jones, D. M. (2000). Effects of hyperlinks on navigation in virtual environments. *International Journal of Human Computer Studies*, *53*(4), 551–581. http://doi.org/10.1006/ijhc.2000.0402

Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Humam Learning and Memory*, *6*(5), 516–528.

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. http://doi.org/10.1002/hipo.22111

Stern, E., & Leiser, D. (2010). Levels of Spatial Knowledge and Urban Travel Modeling. *Geographical Analysis*, *20*(2), 140–155. http://doi.org/10.1111/j.1538-4632.1988.tb00172.x

Strickrodt, M., Bülthoff, H. H., & Meilinger, T. (2018). Memory for Navigable Space is Flexible and Not Restricted to Exclusive Local or Global Memory Units. *Journal of Experimental Psychology: Learning , Memory , and Cognition*.

Suma, E. A., Lipps, Z., Finkelstein, S., Krum, D. M., & Bolas, M. (2012). Impossible Spaces: Maximizing Natural Walking in Virtual Environments with Self-Overlapping Architecture. *IEEE Transactions on Visualization and Computer Graphics*, *18*(4), 555–564. Retrieved from http://people.ict.usc.edu/~suma/papers/suma-vr2012.pdf

Tcheang, L., Bulthoff, H. H., & Burgess, N. (2011). Visual influence on path integration in darkness indicates a multimodal representation of large-scale space. *Proceedings of the National Academy of Sciences*, *108*(3), 1152–1157. http://doi.org/10.1073/pnas.1011843108

Tlauka, M., Carter, P., Mahlberg, T., & Wilson, P. N. (2011). The first-perspective alignment effect: The role of environmental complexity and familiarity with surroundings. *The Quarterly Journal of Experimental Psychology*, *64*(11), 2236–2250. http://doi.org/10.1080/17470218.2011.586710

Tolman, E. C. (1948). Cognitive maps in rats and men. *The Psychological Review*, *55*(4), 189–208. Retrieved from http://psychclassics.yorku.ca/Tolman/Maps/maps.htm

Tversky, B. (1981). Distortions in Memory for Maps. *COGNITIVE PSYCHOLOGY*, *13*, 407–433. Retrieved from https://www.tc.columbia.edu/faculty/bt2158/faculty-profile/files/1981_Tversky_Distortionsinmemoryformaps.pdf

Vasylevska, K., Kaufmann, H., Bolas, M., & Suma, E. A. (2013). Flexible Spaces: Dynamic Layout Generation for Infinite Walking in Virtual Environments. *In Proceedings of the IEEE Symposium on 3D User Interfaces* (*3DUI)*, 39–42.

Wang, R. F. (2016). Building a cognitive map by assembling multiple path integration systems. *Psychonomic Bulletin & Review*, *23*(3), 692–702. http://doi.org/10.3758/s13423-015-0952-y

Wang, R. F., & Brockmole, J. R. (2003a). Human navigation in nested environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 398–404. http://doi.org/10.1037/0278-7393.29.3.398

Wang, R. F., & Brockmole, J. R. (2003b). Simultaneous spatial updating in nested environments. *Psychonomic Bulletin & Review*, *10*(4), 981–986. http://doi.org/10.3758/BF03196562

Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, *77*(3), 215–250. http://doi.org/10.1016/S0010-0277(00)00105-0

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. http://doi.org/10.1016/j.cognition.2017.05.020

Wehner, R., & Srinivasan, M. V. (1981). Searching behaviour of desert ants, genus Cataglyphis (Formicidae, Hymenoptera). *Journal of Comparative Physiology A*, *142*(3), 315–338. http://doi.org/10.1007/BF00605445

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. http://doi.org/10.1037/a0035261

Wilson, P. N., & Wildbur, D. J. (2004). First-perspective alignment effects in a computer-simulated environment. *British Journal of Psychology*, *95*, 197–217. http://doi.org/10.1348/000712604773952421

Wilson, P. N., Wilson, D. a, Griffiths, L., & Fox, S. (2007). First-perspective spatial alignment effects from real-world exploration. *Memory & Cognition*, *35*(6), 1432–44. http://doi.org/10.3758/BF03193613

Zetzsche, C., Wolter, J., Galbraith, C., & Schill, K. (2009). Representation of space: Image-like or sensorimotor? *Spatial Vision*, *22*(5), 409–424. http://doi.org/10.1163/156856809789476074

Zhao, M., & Warren, W. H. (2015a). Environmental stability modulates the role of path integration in human navigation. *Cognition*, *142*, 96–109. http://doi.org/10.1016/j.cognition.2015.05.008

Zhao, M., & Warren, W. H. (2015b). How You Get There From Here. *Psychological Science*, *26*(6), 915–924. http://doi.org/10.1177/0956797615574952

7.4 Study 4: A hierarchy of reference frames

Study reference

Title: Memory for navigable space is flexible and not restricted to exclusive local or global memory units

Authors: Marianne Strickrodt, Heinrich H. Bülthoff, Tobias Meilinger

Status: Accepted, Published

How to cite: Strickrodt, M., Bülthoff, H.H., & Meilinger, T. (2018). Memory for navigable space is flexible and not restricted to exclusive local or global memory units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Abstract

Objects learned within single enclosed spaces (e.g., rooms) can be represented within a single reference frame. Contrarily, the representation of navigable spaces (multiple interconnected enclosed spaces) is less well understood. In this study we examined different levels of integration within memory (local, regional, global), when learning object locations in navigable space. Participants consecutively learned two distinctive regions of a virtual environment that eventually converged at a common transition point and subsequently solved a pointing task. In Experiment 1 pointing latency increased with increasing corridor distance to the target and additionally when pointing into the other region. Further, when pointing within a region alignment with local and regional reference frames, when pointing across regional boundaries alignment with a global reference frame was found to accelerate pointing. Thus, participants memorized local corridors, clustered corridors into regions and integrated globally across the entire environment. Introducing the transition point at the beginning of learning each region in Experiment 2 caused previous region effects to vanish. Our findings emphasize the importance of locally confined spaces for structuring spatial memory and suggest that the opportunity to integrate novel into existing spatial information early during learning may influence unit formation on the regional level. Further, global representations seem to be consulted only when accessing spatial information beyond regional borders. Our results are inconsistent with conceptions of spatial memory for large scale environments based either exclusively on local reference frames or upon a single reference frame encompassing the whole environment, but rather support hierarchical representation of space.

Keywords: spatial memory; reference frames; hierarchical models of spatial memory; levels of spatial integration

Introduction

When learning a new environment, for example, after moving to a new city, spatial memory of our surrounding is gathered successively. Each time we leave the house we visit multiple places sequentially, thus, we are confronted with chunks of spatial information over time. One day we leave our home and walk to the market place, thereby passing the bank, the pharmacy and the bakery. Yet, on another day, we take the bus or subway to the remote book shop in another borough of the city. From there we stroll through the city, passing a café, then a butcher, until we, unexpectedly, end up at the market place again. Not only our everyday experience but also scientific evidence shows that we are quite able to learn the relationship between two (or more) separately learned neighborhoods. Schinazi, Nardi, Newcombe, Shipley, and Epstein (2013), for example, who had people learn two separate routes first and introduced the connecting route in a second learning session, found a rapid learning of between route constellation of landmarks (see also Ishikawa & Montello, 2006; Weisberg, Schinazi, Newcombe, Shipley, & Epstein, 2014). What remains unclear: How do we deal with those chunks of information? How do we relate the locations of pharmacy and café without having experienced them coincidentally? These questions are addressed within the present work.

There are theories postulating that we rely on a network of locally confined places when representing navigable space, such as buildings or cities (Chrastil & Warren, 2014; Meilinger, 2008; Warren, Rothman, Schnapp, & Ericson, 2017). Chrastil and Warren (2014), for example, postulate the use of a labelled graph. Places are represented as nodes and the connectivity between places are represented as edges. This graph is labelled with additional, metric information, such as distances between the places and angles between the edges starting from the same node. In the graph model of Meilinger (2008) nodes represent reference frames limited to vista spaces, i.e., places one can see from a single vantage point (e.g., a street, a room) (Montello, 1993). Object locations and surrounding geometry within a vista space are stored relative to this local reference frame. The edges connecting local reference frames are specified perspective shifts (distance and angles). In both theories, a complex navigable space is represented without the need for a coarser, global integration that is stored in long-term memory, but with the help of local information only. To point to distant targets from memory local information are integrated during the recall process online within working memory to, for example, form a transient reference frame incorporating one's current position and the target

(Meilinger, 2008). Following these theories new connected routes are simply added to the existent graph representation by adding new nodes and edges.

There is ample evidence that immediately visible surroundings (i.e., vista spaces) are core units when memorizing and dealing with navigable space. Updating of objects located in a room is disrupted when leaving this room (Wang & Brockmole, 2003a). Visual borders clustering an object array seem to likewise cluster the representation of the array (e.g., Kosslyn, Pick, & Fariello, 1974; Marchette, Ryan, & Epstein, 2017; Meilinger, Strickrodt, & Bülthoff, 2016). Also, when participants learn the layout of objects that are spread across multiple streets or corridors (i.e., multiple interconnected vista spaces) evidence was found that multiple reference frames are memorized, which are defined by and confined to the immediately visible surrounding of a street and/or corridor (Meilinger, Riecke, & Bülthoff, 2014; Werner & Schmidt, 1999). Spatial memory based on reference frames is usually detected by utilizing one key characteristic of reference frames, namely, the orientation dependency of spatial memory. Accessing relative directions between object locations that are stored in memory is faster and more accurate when one is (physically or mentally) aligned with certain orientations compared to other orientations. This is explained by encoding objects relative to one or more orthogonal reference axes. Alignment with an axis allows for a rather effortless retrieval, while being misaligned requires costly additional transformations (McNamara, Sluzenski, & Rump, 2008; Mou, McNamara, Valiquette, & Rump, 2004). Meilinger, Riecke and Bülthoff (2014) showed that participants that learned a rather complex virtual environment consisting of seven corridors performed best in a subsequent pointing task when they were bodily aligned with the initial view within each single corridor. This indicates that local memory units were stored, more precisely, in the form of one reference frame for each corridor.

An alternative approach to spatial memory structures based on mere local units are hierarchical conceptions of spatial memory (e.g., Mallot & Basten, 2009; Stevens & Coupe, 1978; Wiener & Mallot, 2003). They do not oppose concepts based on mere local units but extend them. They assume multiple levels of representation and start from the premise that local spatial entities (places) can be subsumed under another memory unit on a superordinate hierarchical level to form a region, for instance. There is evidence that we indeed represent regions and use this information for spatial tasks. Route decisions were shown to be made according to the least number of transitions between predefined regions (e.g., Schick, Halfmann, & Mallot, 2015; Wiener & Mallot, 2003) and judgements about the relative position of two remote cities (e.g., San Diego in

California and Reno in Nevada) was found to be based on the relative position of the two federal states the cities are in (e.g., California is west of Nevada), rather than the actual location of the cities (e.g., San Diego, California is east of Reno, Nevada) (Stevens & Coupe, 1978). Such studies, however, remain uninformative about the format of this superordinate memory unit. What do we store on the higher level of hierarchy? Do we store conceptual or topological knowledge (Hirtle & Mascolo, 1986; Meilinger, 2008; Wang & Brockmole, 2003a, 2003b; Wiener & Mallot, 2003), or do we embed multiple vista spaces that form a region into a new metric relational scheme, another superordinate spatial reference frame (e.g., McNamara, Sluzenski, & Rump, 2008; Meilinger & Vosgerau, 2010)?

In fact, it was shown that reference frames can spread across multiple adjacent vista spaces. For example, Wilson, Wilson, Griffiths and Fox (2007) had people learn object locations within a simple three-corridor environment, containing four target objects. A subsequent pointing task provided evidence for a global reference frame covering all three corridors. Irrespective of the corridor participants were in, they performed best when aligned with the perspective of the very first corridor compared to other orientations. Thus, there seem to be circumstances under which spatial information from multiple enclosed vista spaces are aggregated and integrated into one reference frame (see also Meilinger, Frankenstein, Watanabe, Bülthoff, & Hölscher, 2015; Richardson, Montello, & Hegarty, 1999; Tlauka, Carter, Mahlberg, & Wilson, 2011).

An interesting study by Greenauer and Waller (2010) showed that people seem to build not only micro- but also macroreference frames when learning two arrays of objects within the same room (vista space). Depending on whether participants pointed within one object array or within the other, orthogonally-aligned object array or whether they pointed across arrays, different body perspectives were identified to elicit best pointing performance. In the case of within-array pointing best pointing performance was found when participants were aligned with the geometry (i.e., main axis) of the respective object array pointing was currently performed in. This indicates that two orthogonally aligned microreference frames were formed, one reference frame for each object array, and used when the current position and the target object both are located within the same microreference frame. In contrast, when the participants imagined themselves to be standing within one object array, but the target was located in the other object array another, single orientation was found to elicit best pointing performance. In these across-array pointing trials evidence for a macroreference frame was found that aligned with the salient axis of the between-array geometry.

This indicates that a hierarchy of reference frames can be formed for chunks of spatial information presented in a vista space. For spaces extending a vista space a hierarchy of reference frames was discussed in the literature (McNamara et al., 2008; Meilinger & Vosgerau, 2010) assuming that reference frames for vista spaces can be grouped together to form distinct regions which are integrated to form distinct regional reference frames on a higher level of the hierarchy. However, it was not yet experimentally tested for.

Altogether, past research produced results that indicate that local vista spaces serve as memory units for navigable space (e.g., Marchette, Ryan, & Epstein, 2017; Meilinger, Riecke, & Bülthoff, 2014; Meilinger, Strickrodt, & Bülthoff, 2016), but likewise can be aggregated and integrated into a single reference frame (e.g., Meilinger, Frankenstein, Watanabe, Bülthoff, & Hölscher, 2015; Wilson, Wilson, Griffiths, & Fox, 2007). Additionally, regions that are clustering navigable space are stored and used for route planning (e.g., Wiener & Mallot, 2003) and relational judgements (e.g., Stevens & Coupe, 1978). Therefore, in this study we investigate whether spatial memory acquired in navigable space is stored on multiple levels, similar to the micro- and macroreference frames found in vista spaces (Greenauer & Waller, 2010). To examine this, we built a clustered virtual environment consisting of two obliquely aligned regions. Corridors within a region were similar in many attributes (e.g., color, category of landmarks, distance, see Materials) while being maximally different from the attributes of the other region. Assuming a hierarchy of reference frames opens two questions: First, how many hierarchical levels are formed, and which areas do units on each level comprise? For example, multiple local units might be stored on a single level where each reference frame is limited to a single corridor (vista space). Corridors belonging to the same region might be aggregated and integrated to form a regional memory unit on a superordinate level. It is also possible that a global memory unit is formed that spans across the entire environment, thus, comprising all spatial information from all encountered vista spaces of the environment within a single reference frame. Different combinations of local corridor units, regional units or a global unit are conceivable, for example, local memory units on the lower level and a single, global reference frame on a superordinate level of hierarchy, while there is no hierarchical level for regional units. The second question is: What sets the main orientation of the reference frames on the different levels of hierarchy? We addressed these questions by assessing participants performance in a spatial memory task where they pointed from selected locations in the environment to memorized landmarks.

The first question—expansion of reference frames and number of levels—is addressed based on two main assumptions: (a) Each new memory unit that is accessed during recall will lead to an increase in pointing latency (see also Meilinger, Strickrodt, & Bülthoff, 2016) and (b) being aligned with the orientation of a stored reference frame during recall will lead to fastest pointing latency compared to being aligned otherwise (e.g., McNamara, Sluzenski, & Rump, 2008; Mou, McNamara, Valiquette, & Rump, 2004). We focus on latency rather than accuracy, based on the assumption that both are indicative of distinct processes of learning and memory (e.g., Prinzmetal, McCool, & Park, 2005; Sternberg, 1969). We more strongly associate pointing accuracy with the precision of memory set mainly during the encoding process, whereas latency more strongly relates to the retrieval of the memory content, may it be more or less precise (see also Meilinger, Strickrodt, & Bülthoff, 2016; Pantelides, Kelly, & Avraamides, 2016).

Based on the two above stated assumptions, clear performance patterns were specified that would be in favor of the distinct hierarchical levels. If local corridor units and the successive connections between them are stored, then the relative position of objects will be retrieved via the successive activation of all local memory units that are represented between one's current position in the environment and the location of the target. This means, pointing to a target that is three corridors away requires the consecutive activation of one's current position, the two corridors in-between and the memory unit of the target corridor. Thus, pointing latency should increase with increased corridor distance between current and target corridor, as the number of costly transitions between connected local reference frames increases (corridor distance effect). More precisely, pointing to a target in an adjacent corridor should be faster than pointing to a target two corridors away, and so on. A similar pattern can be predicted for the regional level: If regional units and their connection are memorized, having to point from one's current location to the target in the other region should lead to performance loss as a new memory unit (the other region) must be activated (effect of target region). Both distance effects—corridor distance and target region effect—can occur in parallel. If all corridors are exclusively integrated into a global reference frame covering the whole environment pointing latency should be independent of corridor distance or target region because the memory unit of the environment is activated no matter where people are located. This approach has some similarity with the concept of spatial priming. For example, McNamara (1986) had participants learn the arrangement of multiple objects within a large hall (vista space). The hall was divided into sections by long bands on the floor. He used a primed object recognition task, where the previous trial either contained an object from within the same region (section) or another region, and either far or close with respect to Euclidean distance. One assumption was that recognition time is dependent on the level of activation that was induced by the prime. Another assumption was that recognition priming mirrors an automatic process of retrieving spatial memory, not influenced by recall strategies. Therefore, speed of recognition should tell about the structure of spatial memory. McNamara (1986) found both distance and region effects for this clustered vista space. Our approach is similar to spatial priming as both regard the time needed for recalling spatial memory to be highly dependent and thus informative about the underlying memory structure. Yet it is different as we concentrate on a more complex task in a complex navigable space, the precise recall of target locations during pointing.

The effect of body orientation during pointing should be telling with regards to deployed reference frames. If pointing is based on local reference frames, pointing should be fastest when aligned with the initial view within every corridor. In case of regional reference frames, aggregating multiple corridors, we should be able to identify one main orientation within each region that leads to best performance when aligned with it. Yet again, if a global reference frame is deployed, there should be one facing orientation across the whole environment yielding best pointing performance. All three alignment effects might be observed in parallel.

There are many studies examining reference frame orientation for object locations in vista spaces, but less that examine navigable space. Therefore, as a second question, we ask what sets the reference frame orientation beyond an enclosed vista space, thus, on higher levels of the spatial hierarchy? Vista space studies suggest that there are multiple factors influencing the orientation of a reference system, namely, egocentric experience (e.g., first perspective) (e.g., Kelly & McNamara, 2008; Rieser, 1989), salient layout intrinsic cues (formed rows and columns) (e.g., Kelly & McNamara, 2008), salient layout extrinsic cues (room geometry) (e.g., Shelton & McNamara, 2001; Valiquette & McNamara, 2007) and even instructions (e.g., Mou & McNamara, 2002). Whether all these factors play a role when integrating multiple vista spaces into a reference frame on a higher level is not yet known. However, it seems plausible to assume that the notion of space that is interpreted in terms of a conceptual north (e.g., Shelton & McNamara, 2001) is not only accounting for vista, but also for navigable spaces (e.g., McNamara & Valiquette, 2004).

Some studies examining navigable space emphasize the importance of the first perspective experienced within the environment (e.g., Tlauka, Carter, Mahlberg, & Wilson, 2011; Wilson, Wilson, Griffiths, & Fox, 2007) and report, what they call, the First Perspective Alignment effect. Yet again, as has been shown in vista space studies already, the selection of a reference frame direction is not constrained to the assimilation of later information into an already existing reference frame, but can likewise succeed via accommodation (Piaget & Inhelder, 1969), thus, adapting a representation by integrating the cumulated input to form a reference frame with a main orientation distinct from the initial perspective. When learning an object array in vista space the saliency of an initial view can easily be overwritten by emphasizing specific perspectives without directly experiencing it (Greenauer & Waller, 2010; Mou & McNamara, 2002) or by an alignment with a layout intrinsic or extrinsic cue when examining an object array from a different perspective at a later time (Kelly & McNamara, 2010; Shelton & McNamara, 2001). The possibility to update earlier experienced spatial memory through movement can contribute to that (Wang, 2016). When examining the environmental layouts used in studies evidencing a global reference frame in navigable space (e.g., Meilinger, Frankenstein, Watanabe, Bülthoff, & Hölscher, 2015; Meilinger, Riecke, & Bülthoff, 2014; Richardson, Montello, & Hegarty, 1999; Tlauka, Carter, Mahlberg, & Wilson, 2011; Wilson, Wilson, Griffiths, & Fox, 2007) we identified two alternative factors besides the driving force of the first perspective, that might (partly) explain the global alignment. Often the first corridor/street of the environments used in those studies was also the longest. Thus, the first perspective was confounded with the most frequent perspective experienced during the initial walk through the environment. Additionally, often environments with a simple geometric structure were used, for example, three corridors in an upside-down U-shape. Thus, starting from the initial segment participants turned 90° towards the second segment and 90° thereafter following the same turning direction as before, ending up in a corridor parallel to the first one. Already Wertheimer (1923) identified parallelism (among others) as a key principle of perceptual grouping of 2D elements (Gestalt laws). Indeed, recognizing parallelism on 2D planes seems to be a core component of geometrical understanding in humans (see for example Dehaene, Izard, Pica, & Spelke, 2006; Dillon, Huang, & Spelke, 2013). Also, studies investigating human spatial memory in orientation tasks after disorientation within rooms of various shapes showed sensitivity to environmental geometry (e.g., Hermer & Spelke, 1996, 1994). Humans even have the tendency to remember irregular environments as more parallel and regular than they are (e.g., Byrne, 1979; Moar & Bower, 1983; Tversky, 1981). Thus, when confronted with natural, navigable space, we seem to be prone to search for, even superimpose easy structures to conceptualize space and to anchor our knowledge. We therefore reckon that the overall geometry of such a U-shaped environment—two parallel leg segments which converge to the same orthogonal segment, thereby forming a weak bilateral symmetry—is easy to infer and might serve as a salient cue for setting the reference frame orientation along this U-shape. Of course, when first entering the U-shape its structure is not apparent. Only after traversing through the connector segment and leaving the U-shape from the second parallel leg one might be able to understand the configuration. When reentering the U-shape later in time the shape is recognized (i.e., "I am back at the U-shape") and anticipation of the subsequent corridors is facilitated, therefore, boosting saliency of the orientation when entering the U-shape. Using such a cue requires relating multiple corridors and, therefore, might be identified only after sufficient experience with the environment.

We addressed this issue of multiple salient cues for reference frame orientation by examining three potential reference frame alignments for the two superordinate levels: (1) Alignment with the first perspective within each region and across the overall environment, (2) alignment with the most frequently encountered perspective during the first walk through each region and across the overall environment, and (3) alignment with the salient geometric cue of a U-shape within each region or across the environment. To our knowledge this is the first study testing for multiple potential reference frames on three levels of hierarchy. Predictions are described and visualized in the result section of Experiment 1 (Deployed Reference Frames).

In summary, our study aimed to explore the architecture of spatial memory acquired in navigable space. Participants were asked to explore and learn the layout of a virtual environment consisting of two regions by walking. Two main questions were addressed: Do people memorize the environment in the form of local, regional or global memory units or within multiple forms, thus, revealing hierarchical structure of spatial memory? And what sets the main orientation of the reference frames on the different levels of hierarchy, the first perspective, the most frequently experienced perspective, or the geometric salience of a U-shape? Between the two experiments we manipulated the regional start point of learning to vary ease of global integration. In Experiment 1 participants started at the outer ends of both regions and reached the regional transition point only after walking through the entire region. Conceivably, this makes it harder to integrate both regions into a global memory unit, while facilitating regional clustering. In Experiment 2 participants started learning from the transition point between the two regions, enabling them to encode the connection of the two regions from the start and integrate each additional corridor into their existing knowledge. Compared to Experiment 1 this learning procedure might facilitate global integration. After learning the environment, we tested participants spatial memory with a pointing task. Their performance was analyzed for distance and alignment effects on the local, regional and global level to shed light on how their memory was structured.

Experiment 1

Method

Participants

After providing informed consent, 23 participants partook in the experiment, all naïve to the research question and the experiment. They received monetary compensation for participating. The experiment was approved by the local ethics committee. Participants were required to have normal or corrected to normal visual acuity. For three participants, the experiment was terminated before completion: one was unable to reach the learning criteria (see Procedure), two did not finish the experiment in the maximally available time of three and a half hours. We excluded another two because of bad performance (see Results). The remaining 18 subjects were included in the analysis, 6 of which were male. Their average age was 29.72 years (SD = 10.61).

Material

During learning participants walked freely in a large tracking space of 12×12 m, while their head position was tracked by 16 high-speed infrared cameras at 150 Hz (Vicon® MX 13) and transmitted to a computer (with NVIDIA GTX1080 graphics card). The computer rendered the respective egocentric view within the virtual environment, which was displayed to the participants via an Oculus Rift Development Kit 2 that provided a field of view of ca. 100°×100° at a resolution of 960×1080 pixels for each eye. Head-mounted display (HMD) and computer were connected via a long cable. The experimenter followed participants closely during learning, carrying and repositioning the cable to ensure safe movement through the tracking space and to eliminate direction cues through the cable origin. This immersive VR setup allowed for a realistic learning experience, including proprioceptive cues and stereopsis as well as self-determined exploration of the virtual world.

The virtual environment (Figure 1) consisted of eight interconnected corridors, each containing one distinct virtual 3D object standing on a pedestal in the middle of the corridor on either side. The environment was constructed to induce maximal separation between the two regions of the environment. Besides visual cues (colored walls and pedestals: blue and red) additional factors were held constant within one half while being maximally different from the other region: semantic similarity (categories: animals used as landmark objects in the blue region, red region contained only tools), distance between regions (longest corridors in the middle of maze at regional transition point), and complexity of turning angles between corridors (90° angles within a region, 45° at transition point). Also, the two halves of the environment were explored one after another (spatio-temporal learning contingency), with an overlap only at the regional transition point (mandala). Each region comprised the same number of corridors and objects, ensuring similar memory load.

The objects served as landmarks for self-localization and orientation and as target objects in a subsequent pointing task. For each participant, a sample of eight landmarks (four animals, four tools) were randomly selected from a pool of 16 objects (eight animals, eight tools) and distributed across corridors (animals in blue, tools in red corridors). The remaining eight objects served as distractors objects in a primed recognition task (not described here). The pool of 16 objects had previously been validated in a pre-test and selected from a set of 24 objects as being well recognizable. In the virtual environment landmarks were either standing on the right or on the left corridor wall and facing either direction along the corridor. For each participant, the four possible combinations of side and facing (four possible positions of a landmark) were assigned randomly to the four corridors of a region.

Procedure

Participants were instructed that they were about to learn a virtual environment, starting with one half of the environment, then followed by the other half of the environment. Participants either started in the blue or in the red region (pseudo-randomly assigned) to account for saliency effects of individual regions.

Learning phase Before exploring the environment participants were briefed about the experimental procedure. They were informed that they were about to learn a virtual environment, first one then the other half of the environment and that they need to memorize the whole environment and how the objects and the corridors relate to each other. The instructions made clear that they will be teleported to different positions within the environment later, where they must be able to orient themselves and to relate to any other object from there. Participants then were equipped with the HMD. With the display turned off, they were disoriented, led through a few left and right turns to a predefined starting position.

In both halves participants started at the very end of a region, standing with their back to the dead-end of the last corridor, facing along the corridor (end-to-transition condition). They then walked through the corridors towards the transition point (mandala). Standing at the transition point the view into the other, yet unexplored region was blocked by a curtain. Only parts of the corridor of the other region was visible, however, neither the object and the pedestal nor the turn to the next corridor of the other region was visible. Standing at the mandala they had clear visual information about how the two corridors of the two regions are connected to each other. Here participants were explicitly informed: “You are now at the point where the two halves of the environment come together. Only currently the view into the next corridor is blocked. You will learn this part of the environment after you have learned the first half”. The experimenter emphasized again that participants would be required to relate objects from both halves of the environment later in the testing phase and therefore would need to know how the halves relate to each other.

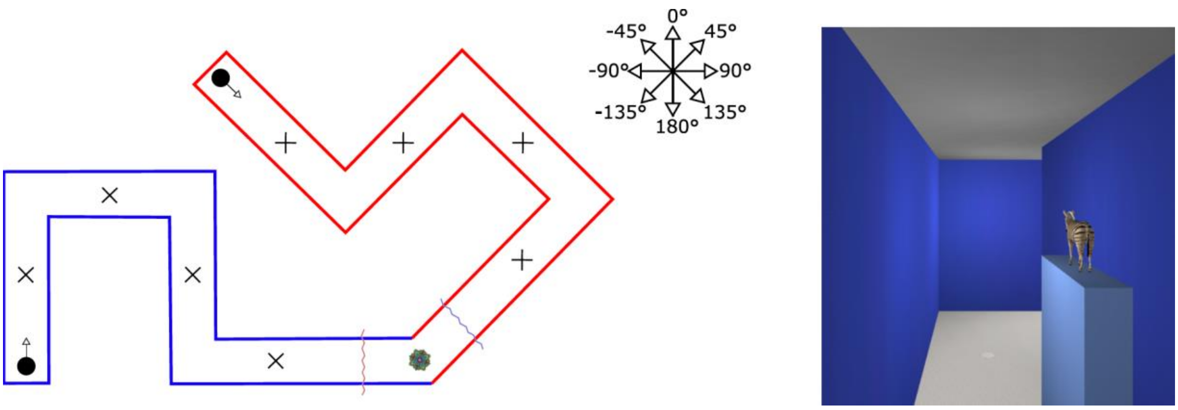


Figure 1. Left: Schematic, aerial view of the environment. Each corridor contained one object (x). A mandala marks the transition point between the two regions. Black dots with arrows indicate initial positions and orientation within both regions (Experiment 1). Wave lines indicate the approximate position of the curtain obstructing the view into the other region when standing at the mandala. Star of black arrows: Eight body orientations tested during pointing while standing at different predefined positions within the environment (at the x’s within each corridor). Right: Example of egocentric view in the environment as experienced by participants. See the online article for the color version of this figure.

Participants walked from the dead-end to the mandala and back again to the dead-end corridor three times. They had to name the objects during learning to ensure they could identify each object by name. Apart from that, participant could walk in their own speed and stop anytime they liked to look around the environment. However, they were not allowed to walk back before fully translating the region.

After learning the environment three times a learn-check was carried out. Objects and pedestals were removed from the scene and fog occluded the turning direction to the next corridor. While walking through the deprived environment, participants had to recall from memory name, side, and facing direction of the object in the respective corridor as well as the turning direction to get into the next corridor. All four corridors were queried, two from each learning direction (dead-end to transition point and vice versa). If participants did not reach 100% accuracy (learning criteria), they had to learn the environment again. This time they walked to the mandala and back only once and then did the learn-check again. This procedure was repeated until participants reached the learning criteria, or a maximum of six times. Participants were excluded from the experiment in case they were not able to learn the environment within a maximum of six learning repetitions. After reaching the learning criteria participants took a short break and continued learning the second half of the environment, following the same end-to-transition point procedure as in the first region. When reaching the transition point for the first time in the second region they were explicitly informed that "This is where the two parts of the environment come together. You are now at the exact same position as before when you were at the end of the first half of the environment.", and the experimenter emphasized again that it is important to know how the two halves relate to each other to solve the subsequent memory test. Like this, we ensured participant would pay enough attention to this part of the environment, a part that they would never fully explore by walking across this point. The learning phase ended after reaching the learning criteria for the second region as well. After a short break, the testing phase started.

Testing phase The first task was a primed recognition task, which is not reported here. Participants recognized target objects from the environment among distractors after being primed by other target objects or distractors. The task took approximately 20 minutes. Subsequently, after a short break, the pointing task started.

During pointing participants sat on a high chair. A joystick was used for recording responses, standing right in front of the participant at belly height.

In each trial of the pointing task, participants were teleported to the middle of a corridor of the environment next to a landmark. Their head and body was aligned with one of eight possible orientations, evenly spread around 360° in steps of 45° (Figure 1, black arrows). First, they looked around, self-localizing themselves with the help of the object in the corridor. On both sides of the corridor arms the view was blocked by a white fog, thus, the turning direction to the next corridor was not visible. After confirming self-localization and orientation by clicking a button on the joystick, the name of one of the remaining seven objects was projected over the rendered environment and participants were required to point to this object. This means, they had to indicate the direction towards the target with respect to their current position and orientation. The direction had to be indicated by moving the joystick handle (allowing for the full range of 360°) and confirming the target direction with a button press. No visual feedback about the joystick direction was given. Participants had to look straight ahead while giving their response (both for self-localization and pointing) otherwise their response was not registered. Afterwards, the experiment continued with the next trial.

From each of the eight corridors participants pointed eight times, covering all eight body orientations, to a target within the same region, and again eight times, covering all body orientations, while pointing to a target of the other region. This adds up to 128 trials. Target objects were selected randomly for every trial under the constraint that each object within a region was used equally often as a target. Trial order was random. Each task started with eight practice trials, randomly chosen from the pool of 128 trials, to familiarize participants with the task. Those were not included in the analysis. Including the practice trials there were 136 trials with four blocks of 34 trials each. Each block was followed by a pause of self-determined length. Participants were instructed to point as fast and as accurate as possible. No feedback was given about the accuracy of the response and task execution was not constrained by any time limit. At the end of the experiment we asked participants to judge their general sense of direction (SOD) on a five-point Likert scale.

To summarize, we varied the within-subject factors body orientation (while pointing being aligned with 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315° relative to the environment, see Figure 1), target region (target within same region vs. across region with respect to current position) and corridor distance (current position and target one to seven corridors apart). We recorded self-localization time (not reported), pointing latency (i.e., the time between displaying the target name and registering a response), and pointing error (i.e., absolute deviation between correct and indicated target direction).

Results

As mentioned in the sample description we excluded two participants based on their pointing performance. Either pointing accuracy did not significantly exceed chance level or latency was 2 *SD* slower than the sample mean. We further excluded data points deviating more than ±2 *SD* from a participant's overall mean pointing latency and pointing error[18]. On average, we excluded 4.12% (*SD* = 2.57) values from pointing error and 4.64% (*SD* = 1.33) values from pointing latency per participant. Participants needed on average 2.00 learning repetitions (*SD* = 1.19) to learn the blue region (reflecting about 8 walks through the environment, exclusive of walks during learn-check) and 2.11 learning repetitions (*SD* = 1.13) to learn the red region (reflecting about 8.22 walks). They spend on average 35.79 (*SD* = 12.07) minutes in the environment.

Distance and cluster effects

We conducted an ANOVA with the within-subject factors distance and target region. Across region pointing trials with a distance to the targets of more than three corridors were excluded from the analysis to match corridor distance for both target locations (within and across region). Thus, 88 out of the 128 trials were used for this analysis, 64 within-region and 24 across region trials.

Pointing latency (Figure 2, left) was indicative of mental separation between corridors and regions. We found main effects of distance, $F(2, 34) = 9.74$, $p < .001$, $\eta_p^2 = .36$, and target region, $F(1, 17) = 4.88$, $p = .041$, $\eta_p^2 = .22$. Pointing latency significantly differed between corridor distance 2 and 3, $t(34) = -2.856$, $p = .019$, as well as 1 and 3, $t(34) = -4.342$, $p < .001$, but not between corridor distance 1 and 2, $t(34) = -1.487$, $p = .310$ (p-value adjusted by Tukey method for multiple comparisons). The response was on average 1.98 sec faster when pointing within the same region compared to pointing to a target in the opposite region. No interaction between target region x distance was found,

[18] Exclusion of latency values did not coincide with exclusion of error values of the same trial. Rather outlier exclusion was done separately for both dependent variables. As stated in the instruction we reckon latency and error to mirror different aspects of spatial memory (precision of memory vs. retrieval of the memory content) (Meilinger, Strickrodt, & Bülthoff, 2016; Pantelides, Kelly, & Avraamides, 2016). Mutual exclusion of latency and error values (i.e., excluding the whole trial) did lead to higher percentage of excluded trials per participant, but did not lead to any significant changes to the results of separate exclusion, which are reported in this paper.

F(1.41, 23.92) = 2.19, *p* = .146, $\eta_p{}^2$ = .11 (Greenhouse-Geisser-adjusted). Thus, there is no evidence that the effect of target region varies across distances.

For pointing error (Figure 2, right) we found a main effect of distance, *F*(2, 34) = 19.57, *p* < .001, $\eta_p{}^2$ = .54. Pointing error was significantly lower for distance 1 compared to distance 2, *t*(34) = -3.978, *p* = .001, as well as compared to distance 3, *t*(34) = -6.171, *p* < .001, but not significantly different between distance 2 and 3, *t*(34) = -2.193, *p* = .087 (p-value adjusted by Tukey method for multiple comparisons). We did not observe a main effect of target region, *F*(1, 17) = 0.35, *p* = .561, $\eta_p{}^2$ = .02, or an interaction of target region x distance, *F*(2, 34) = 2.44, *p* = .102, $\eta_p{}^2$ = .13. Thus, the analysis does not provide evidence that pointing error is higher when pointing across regional boundaries as compared to pointing within ones' current region.

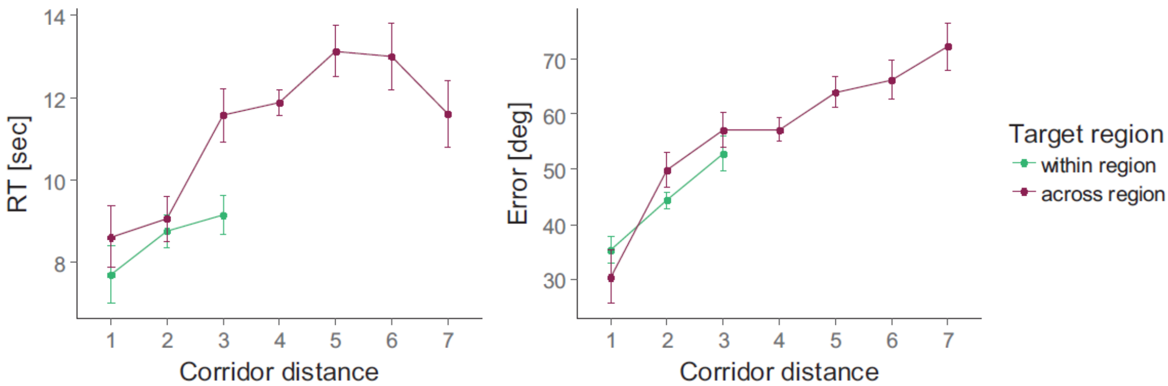


Figure 2. Distance and cluster effects in Experiment 1. Pointing latency (left) and error (right) as a function of corridor distance to the target, pointing to targets within the same region versus across-region. Means and SEMs are depicted. Only data points from corridor distances 1, 2, and 3 were analyzed. See the online article for the color version of this figure.

Adding gender, first learned region (blue or red), number of learning trials, and self-reported sense of direction (SOD) as covariates to the analyses of latency and accuracy did not change the inference statistical values and interpretation of the abovementioned effects in a meaningful way. The results described here are therefore limited to reports of statistics without the consideration of covariates. Further, no main effect of gender, first learned region (blue or red), and SOD on latency or accuracy were found. Only the number of learning trials was associated with pointing error, *r* = 0.69, *p* = .002. Participants who quickly learned the environment were also comparatively better pointers.

Deployed reference frames

The further analysis concentrated on the pointing latency to assess the processes underlying spatial recall (compare to Meilinger, Strickrodt, & Bülthoff, 2016; Pantelides, Kelly, & Avraamides, 2016, and see Introduction). In contrast to the distance analysis, where trials with distances beyond three corridors were excluded, we now used the full data set of 128 trials (excl. outlier values) per participant. To account for the strong distance effect we first conducted an ANOVA with distance as the within-subject factor on the full dataset. The computed residual values instead of the raw values of pointing latency were now used for the following analysis of reference frames. By controlling for the known effect of distance on latency, we hereby removed unwanted variance, and in the following, were aiming to explain variance in the data not yet explained by distance. Interpreting residual values is not as intuitive as interpreting raw pointing latency values (e.g., they center on 0), nevertheless they do preserve the relation and magnitude of difference between data points that are manifest in the raw data (i.e., faster or slower pointing performance). In the following description of the results we will use the more intuitive term of "pointing latency" and "reaction time", but we mean to refer to "latency residuals" after statistically accounting for the effect of distance.

The environment used in the current study consisted of two obliquely aligned regions. Literature suggests that one's current position and target location affect which reference frames are selected to base spatial recall upon. For example, results from Greenauer and Waller (2010) suggest that macroreference frames are used when pointing from one to another object array in a vista space, but not when pointing within one array (see also Zhang, Mou, McNamara, & Wang, 2014, for nested spaces). Therefore, and based on the region effect found in the previous analysis we analyzed the effect of body orientation on pointing performance separately for within and across region trials.

At each position in the environment participants were tested while being aligned with one of eight body orientations.[19] Depending on the reference frame(s) deployed clear predictions can be made about which body orientation should elicit fastest memory access compared to the other body orientations. Figure 3 illustrates our predictions. For each prediction at each position in the environment one orientation can be identified as being superior. If participants use local reference frames, pointing latency should be fastest when being aligned with the first perspective experienced in each corridor compared to be

[19] Supplementary material, Figure S1 and Figure S2, left, give an overview of the pattern of pointing latency before further aggregation to test for the specific reference frame predictions. The figures provide a visualization of pointing latency as a function of body orientation on corridor scale (Figure S1, left) and on global scale (Figure S2, left) (global scale corresponds to compass rose in Figure 3).

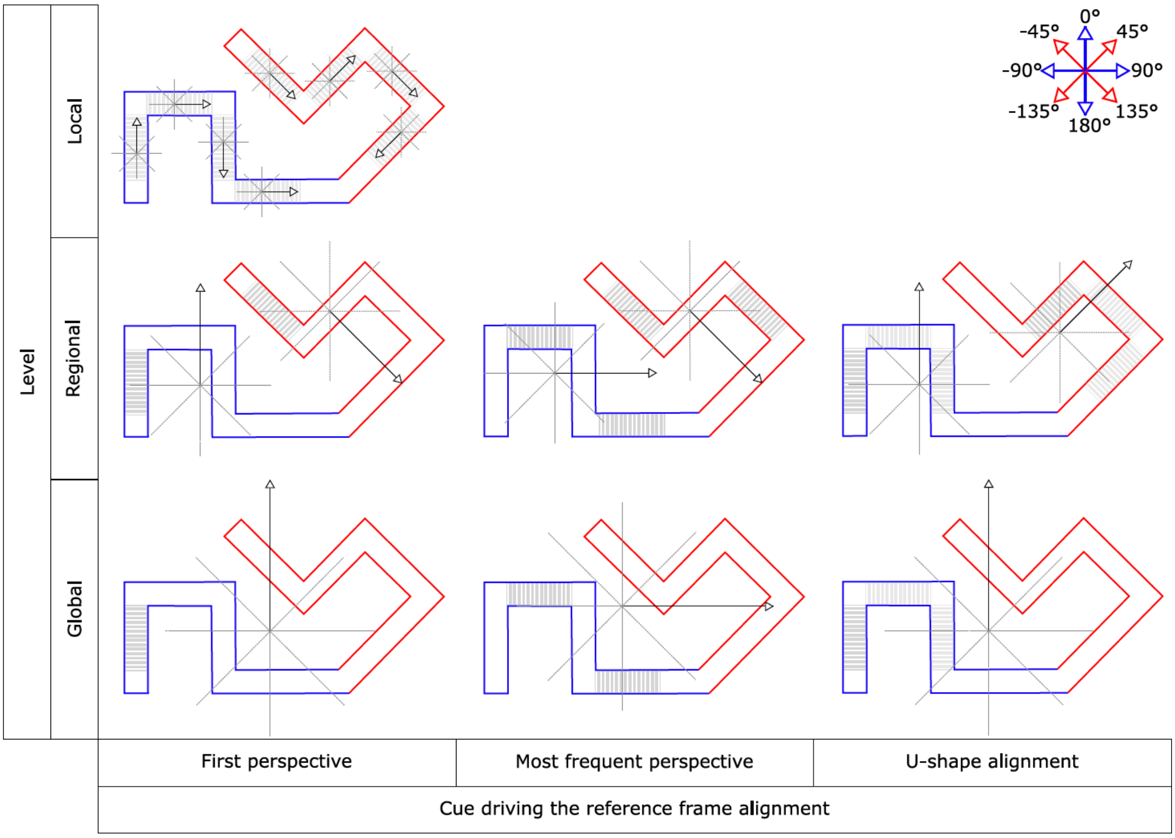


Figure 3. Visualization of the predictions for formed reference frames. We tested for reference frames on the local, regional, or global level (rows) that are either following the first perspective, the most frequent perspective or the alignment of the U-shape of three consecutive corridors (columns). Cues driving the reference frames are marked with gray lines in the respective corridor(s). Following each prediction one orientation can be identified as being superior, marked by a black, solid arrow, compared with the remaining orientations, marked by gray, dashed lines. Being aligned with this orientation should yield fastest spatial recall compared to being aligned otherwise. The predictions depicted here are exemplary—they represent predictions for Experiment 1 (learning from dead-end to transition point). Further, the global prediction accounts for subjects that start off learning the blue region and later continue with the red, therefore, global reference frame predictions follow the blue (first learned) region. Learning from transition point to dead-end (Experiment 2) and starting off with region red (half the sample of Experiment 1 and 2) are not depicted here but can be inferred following the same logic. See the online article for the color version of this figure.

ing aligned with one of the seven remaining orientations (e.g., facing a corridor wall). If participants use a reference frame covering a whole region and the reference frame direction was set by the first perspective experienced in this region, latency should be fastest when being aligned with the perspective experienced in the very first corridor, independent of the current position within this region. Thus, for example, facing the left wall in the second corridor of the blue region, should yield faster response times than being aligned otherwise (e.g., facing straight along the corridor). Likewise, if participants form and use a global reference frame covering the whole environment and the reference frame direction was set by the first perspective experienced in the first region, pointing latency should be fastest when being aligned with this first perspective at each single position tested. This means, having started off learning the blue region followed by the red region, should yield faster pointing when being aligned with the first experienced orientation in the blue region – even if located in the red region during recall. In this latter case the superior orientation is obliquely aligned with one's current visual scene in a red corridor. Both on the regional and global scale alternative cues could be used to set the reference frame orientation. Therefore, the most experienced perspective during the first walk through the environment and the salient geometry of the environment, in our case the U-shape of a region are explored as well[20]. For each participant and each prediction, we contrasted the mean pointing latency in the superior orientation trials with the remaining, non-aligned trials. This reaction time differences were then further assessed. A mean positive reaction time difference is indicative of a fit to the respective prediction. Zero corresponds to no difference in pointing latency.

For within region pointing Figure 4, left, side shows the fit for the seven considered predictions, Table 1, left side, the results of the analysis. For each prediction, we analyzed with a t-test whether the reaction time difference is significantly larger than zero. When subjects pointing within their current region we found a significant fit for the prediction of local reference frames and of regional reference frames, both for the first perspective experienced within a region as well as the U-shape. No evidence could be found that participants used a regional reference frame following the most frequently experienced perspective or that they use global reference frames. As mentioned in footnote 2, our reference frame predictions are not uncorrelated. Furthermore, their correlation varies. Local corridor alignment correlates with the prediction of regional reference frame by $r = 0.29$ for first perspective alignment, $r = 0.43$ for most frequently experienced perspective and $r = 0.14$ for U-shape. We therefore reanalyzed the fit to our regional and global reference frame predictions after correcting for the effect of local alignment. Similar results were obtained. We

[20] The predictions tested for are by no means comprehensive, but based on observations of environmental layouts of previous studies. Further, it should be noted that the predictions are not orthogonal. For example, the prediction of a regional reference frame following the first perspective alignment is identical with the prediction of a regional reference frame following the U-shape when considering the blue region, but not when considering the red region (compare to Figure 3). When combining both regions, as done in our analysis, a correlation of first perspective and U-shape prediction remains with $r = .43$. Same holds for the global level, when considering all subjects, half of which learned in the blue, half in the red region. Likewise, the prediction of a regional reference frame following the U-shape is correlated with the prediction of a global reference frame following the U-shape of the first region by $r = .43$. Predictions of U-shape and most frequent alignment can be considered orthogonal with a correlation of $r = -.14$.

still find the fit to the predictions of regional reference frames following the first perspective, $t(17) = 1.785$, $p = .046$, $d = 0.42$, as well as the regional reference frames following the U-shape, $t(17) = 2.309$, $p = .017$, $d = 0.54$, while no other predictions fit well, *t's* < 1.043, *p's* > .155. Similarly, we also corrected the data for the two meaningful regional reference frame predictions separately to see whether we still find a significant fit to the local prediction. The fit to the local prediction remains significant after correcting for the regional U-shape prediction, $t(17) = 1.761$, $p = .048$, $d = 0.42$, and remains as a trend after correcting for the regional first perspective prediction, $t(17) = 1.419$, $p = .087$, $d = 0.33$.

Table 1. T-tests exploring whether the difference in pointing latency of Experiment 1 is significantly larger than zero, which indicates a fit to the respective prediction. For pointing across region (right) only a subset of trials with a maximum corridor distance of three are included in the analysis reported in this table.

		Pointing within region			Pointing across region		
Prediction		*t*(17)	*p*	*d*	*t*(17)	*p*	*d*
Local	First perspective in corridor	2.019	.030 *	0.48	-0.511	.692	-0.12
Region	First perspective	2.176	.022 *	0.51	0.977	.171	0.23
	Most frequent perspective	-0.718	.759	-0.17	-0.170	.566	-0.04
	U-shape geometry	2.622	.009 *	0.62	0.643	.265	0.16
Global	First perspective	0.639	.266	0.15	2.128	.024 *	0.51
	Most frequent perspective	-1.554	.931	-0.37	0.600	.278	0.15
	U-shape geometry	1.039	.157	0.24	2.147	.023 *	0.51

Note. For pointing across region (right) only a subset of trials with a maximum corridor distance of three are included in the analysis reported in this table.
** <.05, one-sided, larger than zero, no correction for multiple comparisons.*

A similar analysis as for within region pointing was conducted for trials in which participants pointed across the regional boundaries. This time, results are inconclusive. Across region data fitted to none of our predictions, *t's* < 1.14, *p's* > 0.136, thus, being aligned with the local corridor or with one of the three possible regional or global reference frame does not seem to facilitate spatial memory access compared to being otherwise aligned. Reckoning that this might be due to greater noise in the across region data, we compared individual standard deviations for within and across region pointing. Indeed, standard deviations were significantly higher in across region trials, $t(17) = -2.25$, $p = .038$, $d = -0.75$, thus, potentially obstruct detection of used reference frames. Therefore, we decided to re-run the same analysis of the fit to the seven predicted reference frames with only a subset of the across region trials, namely, only for trials querying targets that are lying one, two or three corridors away, excluding trials with corridor distance four and higher. With this trial subset (24 trials per participant) the same corridor distances as for the evaluation of the within-region pointing trials are covered, similar as for the analysis of distance and cluster effects. Indeed, individual standard deviation of this subset was comparable to variance for within region trials, $t(17) = -1.50$, $p = .151$, $d = -0.50$. Table 1, right, shows the results of the analysis of across region trials with corridor distance one to three, in Figure 4, right side, latency differences are depicted. As before, data did not show significant fit to neither local nor regional reference frame predictions. However, now significant fits to the predictions of global reference frames were found, more precisely, for the prediction of the first experienced perspective of the first learned region as well as the U-shape experienced in the first learned region. The prediction of a global reference frame based upon the most frequently experienced perspective in the first learned region did not reach significance.

Continuing with the previously used data set we also examined whether pointing across region led to fastest pointing when aligned with the main reference direction of the target rather than the main reference direction of one's current position – a result found by Zhang, Mou, McNamara and Wang (2014) for pointing between two nested spaces. In our study no difference in pointing latency on across region trials could be found between being aligned with the four main orientations of one's current region and being aligned with the four main orientations of the neighboring region containing the target, $t(17) = 0.43$, $p = .674$, $d = .14$, thus, we did not find evidence for the alignment with the target reference frame.

Discussion

In Experiment 1 participants memorized a regionalized environment. They explored both halves of the environment separately by physical walking, starting from the outer ends and walking towards the transition point, where they had a glance into part of the other region. We were interested in how their memory might be structured and what spatial units they form and use to solve a subsequent pointing task. During this pointing task participants were randomly teleported to different locations within the environment, being aligned with one of eight possible body orientations, while pointing to targets within the same region as their current standpoint or to targets located in the other region. The first experiment revealed evidence that participant formed local, regional as well as global spatial units that are accessed during recall.

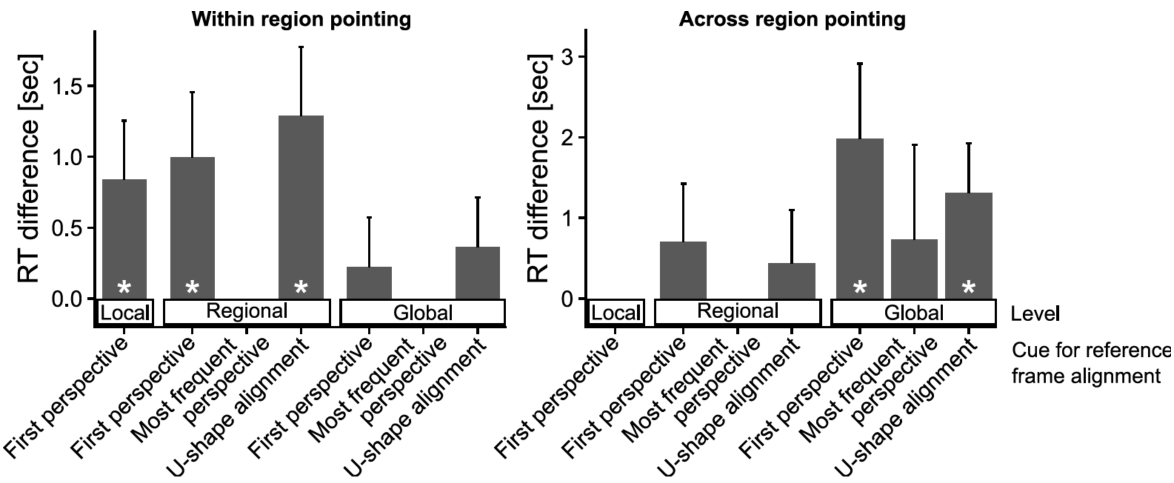


Figure 4. Reference frame alignment in Experiment 1. Difference in pointing latency when being aligned with the superior orientation identified for each prediction compared with the remaining seven orientations, separately for within (left) and across-region trials (right) covering only corridor distance 1, 2, and 3. Positive values are in favor of the prediction, showing faster pointing when aligned with the superior orientation. We do not show negative values in this figure. We tested for seven potential outcomes: local reference frames, and then a reference frame following the first experienced perspective, the most frequently experienced perspective or the salient geometry of a U-shaped environment, either on a regional or a global level. * $p \leq .05$, one-sided, larger than zero, no correction for multiple comparisons.

Pointing latency increased with increasing corridor distance to the target, independent of whether participants pointed within their current region or across regional boundaries. One possible explanation is, that local corridor units were stored in memory which then had to be successively activated in a fixed, learned order, to recall the location of a target object multiple corridors away (e.g., Meilinger, Strickrodt, & Bülthoff, 2016). For every new unit activated the mental effort increases reaction time (and is accumulating error). Besides this latency increase across spatial corridor units, we also found a facilitative effect of local alignment with the corridor. Participants were significantly faster when aligned with the first perspective experienced within each corridor, compared to the seven remaining orientations.

Pointing to targets within the other region, i.e., standing in the red region pointing to a target in the blue region or standing in the blue region pointing to a target in the red region, led to higher pointing latency compared to pointing to a target within one's current region. This effect was not dependent on the corridor distance to the target (no interaction of target region and distance). Additionally, pointing performance in within-region trials was significantly faster when subjects were aligned with the first perspective experienced within a region (first corridor orientation). For example, facing the right wall in the second corridor of the red region, should yield faster response times than being aligned otherwise (except when being aligned with the first experienced perspective in the corridor). Also, pointing performance was significantly faster when subjects were aligned with the orientation when entering the U-shape that is made up by three consecutive corridors. For example, facing the left wall in the third corridor of the red region, should yield faster response times than being aligned otherwise. No fit (even a negative value) was found to the prediction of a regional reference frame following the most frequently experienced perspective within a region. The fit to a regional reference frame following the first perspective and the U-shape remains significant even after statistically accounting for the effect of local corridor alignment. Taken together, participants seem to have formed obliquely aligned, regional reference frames of the blue and red region.

The fit to regional reference frame predictions cannot be explained by a facilitative contra- or orthogonal alignment with the reference frames of the local corridors. This would allow for a speeded access as soon as participants are aligned with either of the four main orientations of the local reference frames (0°, ±90°, 180°). In this case we would expect a fit to all three regional predictions. However, we only find a fit to two predictions.

We further found support for the formation of a global reference frame encompassing both regions. Data of within region pointing did not fit to either of the three predictions for global reference frames. However, pointing performance of across region trials querying targets one to three corridors away revealed a fit to the global predictions following the U-shape as well as the first perspective experienced in the first learned region. This support is somewhat weakened by the fact that it is only found for a subset of the across region trials. Analyzing latency patterns of the full data set of across region trials (corrected for the effect of distance) was non-descriptive. Neither prediction for a specific reference frame alignment fit our data, potentially due to the higher variance in pointing performance for higher corridor distances. Nevertheless, taken equal corridor distances, the selection of remembered reference frames seems to be depending on the relationship between one's current position and the target location. This will be elaborated upon more in the General Discussion. An alternative approach based on results obtained by Zhang, Mou, McNamara and Wang (2014), that showed that the reference frame direction of the target is used, did not describe our data well.

Our results imply that participants embedded multiple vista spaces, formed distinct memory units, one for each region, and memorized the whole environment in a global format. Importantly, as indicated by the alignment effect, those memory units seem to be metric relational schemes, linking multiple vista spaces to one superordinate reference frame direction, thus, they are spatial in nature (compare to McNamara, Sluzenski, & Rump, 2008; Meilinger & Vosgerau, 2010) (see also General Discussion).

Being able to identify a fit to the U-shape prediction besides a first perspective alignment on the regional level is a first indication that—also for navigable space—reference frames are formed that do not merely assimilate newly incoming spatial information into an already existing reference system (i.e., assimilate new corridor information into the direction of the first perspective), but that information of multiple corridors are accumulated and accommodated (geometry of three corridors) to form a reference frame with a new, distinct reference direction (Piaget & Inhelder, 1969). This is in line with studies examining object arrays in vista space, which show that egocentric alignment with a salient geometric cue at a later time point during learning can shape the reference frame accordingly (Kelly & McNamara, 2010; Shelton & McNamara, 2001).

The prediction for a reference frame following the U-shape structure was tailored to the orientation experienced when entering the U-shape, walking towards the connector segment of the two parallel corridors. We decided for a single main orientation to equalize predictions for all potential reference frames (one main orientation vs. the remaining orientations) and because we assumed that the moment one enters the simple shape is particularly prominent as (after enough exposure to the environment) anticipation of the two subsequent corridors is facilitated. During learning each participant entered the U-shape of each region at least six times (3x end-to-transition, 3x transition-to-end). Alternatively, the orientation when leaving the U-shape might be of significant saliency as well, as this orientation is accompanied with the emerging realization of the easy structure of the just travelled corridors. To test for this possibility, we analyzed fit of the within region pointing latency to the prediction of a regional reference following the U-shape with the main orientation centered on the orientation when exiting the shape, thus 180° opposite to the originally tested orientation when entering the U-shape. Interestingly, also here a significant fit was found[21]. Based on this we cannot conclude whether the main orientation of a superordinate reference frame is set when entering or exiting the U-shape. Yet, it remains that the simple structure of two parallel leg segments connected to the same orthogonal segment, thereby forming a weak bilateral symmetry, might serve as an important cue for setting the orientation of superordinate reference frames.

Despite taking longer to recall spatial memory for the other region, pointing accuracy was not affected by the regional belonging of the target. Pointing to targets within the other region did not lead to higher error in pointing. This shows that participants were capable to learn the two halves of the environment separately with overlap at only one common position, the transition point. And this, even though the turning angle at the transition point was more complex (45°) than within region angles (90°)—complexity was previously shown to enlarge pointing error (e.g., Moar & Bower, 1983)—and even though participants were only allowed to examine this transition point visually, but never walked across it. This indicates that the regional transition point was memorized just as accurately as the corridor angles within a region.

Results of Experiment 1 indicate that corridors triggered a clustering of spatial memory into vista space units and that we could successfully trigger clustering into regional memory units. Additionally, it seems that integration into a global unit spanning the whole environment occurred, although this was only detectable for smaller corridor distances. All three levels manifested in spatial metric schemata, the reference frames on the local, regional and global level. In Experiment 2 we set out to replicate findings of Experiment 1 and elucidate the indeterminate findings regarding a global memory unit. Participants in Experiment 1 learned both halves separately, always starting at the outer end of each region, walking towards the transition point. This learning procedure makes it particularly difficult to integrate both regions into a global memory unit. After the first half participants memorized the layout of four corridors and the relative position of the four objects within. When starting to learn the second half, relating each new corridor to the previously formed memory unit was not immediately possible. Participants first had to walk through the four corridors of the second region before ending up at the transition point again. Only then were they able to understand the layout of the whole environment. By reaching the regional transition point again they might

[21] This analysis was not mentioned in the result section. Significant fit of pointing latency to the prediction of a U-shape but with the main orientation centered on the orientation experienced when exiting the U-shape, $t(17) = 2.389$, $p = .014$, $d = 0.56$. Significant fit remains even after correcting for local alignment, $t(17) = 2.116$, $p = .025$, $d = 0.50$.

already have formed a first “draft” of the second region that now must be connected to the representation of the first learned region. Integrating both representations into one spatial unit might entail considerable effort. In a second experiment, we wanted to see whether lowering this effort might lead to clearer results regarding global integration. Regions were still learned separately. However, now participants started learning from the transition point, walking towards the outer ends of each region. Like this, global integration, corridor by corridor, should be possible right away.

Experiment 2

Method

Participants

Twenty-one naïve participants partook in the experiment, receiving monetary compensation. The experiment was approved by the local ethics committee. Participants were required to have normal or corrected to normal visual acuity. Two participants had to stop the experiment before completion: one was unable to reach the learning criteria, one experienced motion sickness during testing. One additional participant was excluded due to chance level pointing performance. The remaining 18 subjects (10 males) were included in the analysis. Their average age was 28.22 years (SD = 8.59).

Material

The material and equipment used was identical to Experiment 1.

Procedure

The procedure was identical to Experiment 1, except of the learning direction. While participants in Experiment 1 started to learn each half of the environment at the outer end of the succession of corridors, walking towards the transition point, participants in Experiment 2 started to learn each half with their initial position at the transition point. Thus, from the beginning they were aware of how both regions are connected to each other. Also, starting with the second region it was immediately possible to relate each newly explored corridor to the previously learned corridors.

Results

We excluded data points deviating more than ±2 SD from a participant’s overall mean pointing latency and pointing error. This resulted in exclusion of on average 5.38% (SD = 2.14) values from pointing error and, separately, 4.39% (SD = 1.20) values from pointing latency per participant. Participants needed on average 1.61 learning repetitions (SD = 0.98) to learn the blue region and

1.61 learning repetitions (*SD* = 0.85) to learn the red region (reflecting about 7.22 walks through the environment for each region, exclusive of walks during learn-check). They spend about 35.21 (*SD* = 7.14) minutes in the environment. We rerun the same steps of analysis as ran in Experiment 1.

Distance and cluster effects

We conducted an ANOVA with the within-subject factors distance and target region. Figure 5 depicts pointing latency and error across the full range of distances. But again, across region pointing trials with a distance to the targets of more than three corridors were excluded from the analysis to match corridor distance between conditions (64 within-region and 24 across-region trials remaining).

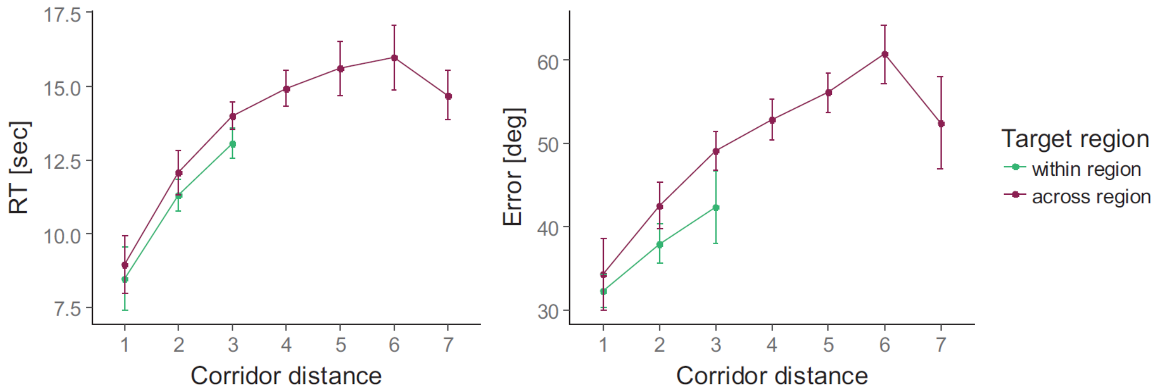


Figure 5. Distance effects in Experiment 2. Pointing latency (left) and error (right) as a function of corridor distance to the target, pointing to targets within the same region versus across-region. Means and SEMs are depicted. Only data points from corridor distances 1, 2, and 3 were analyzed. See the online article for the color version of this figure.

We found a main effect of distance, $F(1.26, 21.34) = 19.01$, $p < .001$, $\eta_p{}^2 = .53$ (Greenhouse-Geisser-adjusted) on pointing latency. Pointing latency differed significantly between corridor distance 1 and 2, $t(34) = -3.762$, $p = .002$, as well as 1 and 3, $t(34) = -6.111$, $p < .001$, but only by trend between corridor distance 2 and 3, $t(34) = -2.349$, $p = .062$ (p-values adjusted by Tukey method for multiple comparisons). In contrast to Experiment 1 no significant effect of target region, $F(1, 17) = 2.06$, $p = .169$, $\eta_p{}^2 = .11$, was found. Thus, now participants didn't need more time to access spatial memory from the other region compared to recalling targets located within the same region. We observed no interaction between target region x distance, $F(2, 34) = 0.15$, $p = .864$, $\eta_p{}^2 = .01$. To further evaluate the non-significant result of target region we run a Bayesian repeated measure ANOVA with the factors distance and target region (r scale for fixed effects = 0.5) to evaluate the likelihood of a null-effect of target region. The highest Bayes factors value was found for the model assuming a single main effect of distance, BF10 = 69,530,000. This is the model that outperformed the null model the most. The model assuming a single main effect of target region revealed a particularly small value, BF10 = 0.371. This is, according to the classification defined by Lee and Wagenmakers (2013) (adjusted from Jeffreys, 1961), within the range of BF10 = 1/3 – 1, which is interpreted as anecdotal evidence for H0 (no effect of target region) compared to H1. Both additionally tested models of two main effects (distance and target region, BF10 = 34,830,000) and the model assuming two main effects and an interaction (BF10 = 4,779,000) only reach BFs smaller than the main effect model of distance. More precisely, the distance model is 2.00 times more likely than the model assuming two main effects and 14.55 times more likely than the full model (main effects and interaction).

For pointing error (Figure 5, right) we again found a main effect of distance, $F(2, 34) = 10.29$, $p < .001$, $\eta_p^2 = .38$. Pointing error was significantly lower for distance 1 compared to distance 2, $t(34) = -2.522$, $p = .042$, as well as compared to distance 3, $t(34) = -4.526$, $p < .001$, but not significantly different between distance 2 and 3, $t(34) = -2.004$, $p = .127$ (p-values adjusted by Tukey method for multiple comparisons). We did not observe a main effect of target region, $F(1, 17) = 1.29$, $p = .272$, $\eta_p^2 = .07$, or an interaction of target region x distance, $F(1.47, 25.00) = 0.49$, $p = .562$, $\eta_p^2 = .03$. Like Experiment 1 participants pointing error was not significantly larger when pointing across regional boundaries compared to pointing within one's current region.

Similar to Experiment 1, adding gender, first learned region (blue or red), number of learning trials, and SOD as covariates to the analyses of latency and accuracy did not change the results for accuracy and latency. We therefore concentrated on reporting statistics without modelling covariates. Besides the number of learning trials, $r = 0.68$, $p = .002$ (more error the more learning trials), no other covariate was associated with latency and error.

Deployed reference frames

After statistically accounting for the effect of distance on pointing latency (see Experiment 1 for a more detailed explanation) we continued with the analysis of deployed reference frames, again concentrating on trials participants pointed within their current region and across regional boundaries separately.[22]

[22] Supplementary material, Figure S1 and Figure S2, right, give an overview of the pattern of pointing latency before further aggregation to test for the specific reference frame predictions. The figures provide a visualization of pointing latency as a function of body orientation on corridor scale (Figure S1, right) and on global scale (Figure S2, right) (global scale corresponds to compass rose in Figure 3).

Figure 6, left, shows the fit for the seven considered predictions for within-region pointing, Table 2, left, the results of the analysis. When participants pointed within their current region we observed a significant fit for the prediction of local reference frames and of regional reference frames, this time for the most frequently experienced perspective. As mentioned in footnote 2 our reference frame predictions are correlated. The strongest correlation can be found for the predictions of local corridor alignment with the regional reference frame following the most frequently experienced perspective ($r = 0.43$). Indeed, when correcting data of within region pointing further for the facilitative effect of local alignment the significant fit to the prediction of regional reference frames following the most frequently exposed perspective disappears, $t(17) = 1.381$, $p = .093$, $d = 0.33$. The fits to the other predictions remain non-significant, *t's* < 0.605, *p's* > .275. Likewise, correcting for the effect of regional alignment with the most frequently experienced perspective, causes the fit to the local reference frame prediction to disappear as well, $t(17) = 0.590$, $p = .281$, $d = 0.14$. The contribution of each single factor hence remains unclear. However, considering the analysis of distance and cluster effects which is supporting local memory units only it is quite likely that local reference frames are producing the observed fits. Data of Experiment 2 did not fit the predictions of regional reference frame following the first perspective or the U-shape. Similar to Experiment 1, no evidence could be found that participants used global reference frames in within region pointing trials.

Finally, trials were analyzed in which participants pointed across the regional boundaries. Again, results for the full set of across region trials are inconclusive. No fit to either prediction of local, regional or global reference frames was found, *t's* < 1.21, *p's* > .122. As again variance in pointing was higher for across compared to within region pointing trials, $t(17) = -2.35$, $p = .031$, $d = -0.78$, we re-ran the reference frame analysis for across region trials with corridor distance one, two and three only (comparable standard deviation for within and across trials, $t(17) = -1.23$, $p = .236$, $d = -0.41$). A figure and results of the analysis can be found in Figure 6, right side, and Table 2, right side. Again, a significant fit to a global reference frame prediction was found, namely, for a global reference frame following the U-shape experienced in the first learned region. Neither of the remaining predictions fitted well. Similar to Experiment 1 participants were not faster when aligned with the reference direction of the target region (four main orientations of the obliquely aligned other region), $t(17) = -1.25$, $p = .230$, $d = -0.42$ (based on dataset with all distances), speaking against an alignment with the target reference frame.

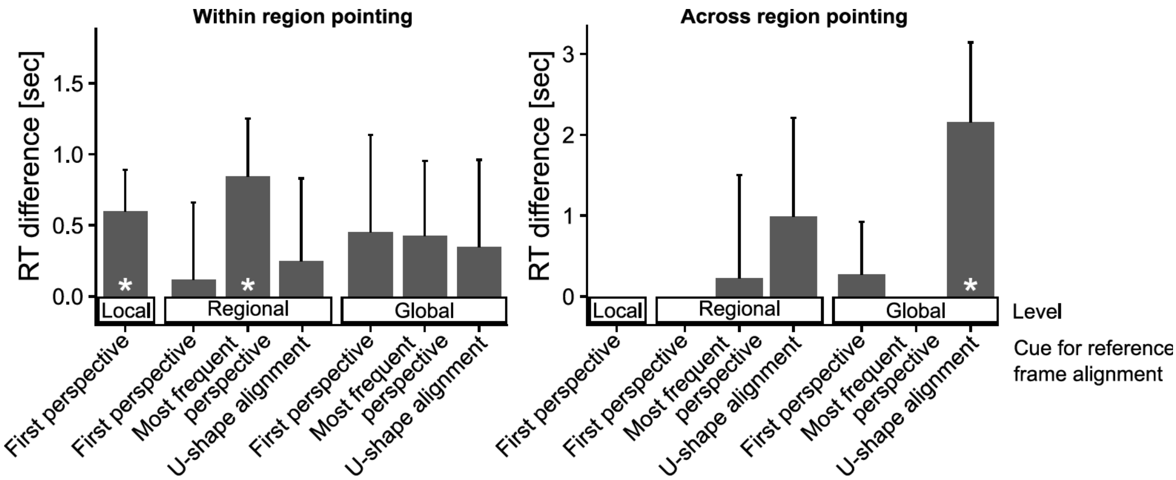


Figure 6. Reference frame alignment in Experiment 2. Difference in pointing latency when being aligned with the superior orientation identified for each prediction compared with the remaining seven orientations, separately for within (left) and across-region trials (right) covering only corridor distances 1, 2, and 3. Positive values are in favor of the prediction, showing faster pointing when aligned with the superior orientation. We do not show negative values in this figure. * $p \leq .05$, one-sided, larger than zero, no correction for multiple comparisons.

Table 2. T-tests exploring whether the difference in pointing latency of Experiment 2 is significantly larger than zero, which indicates a fit to the respective prediction. For pointing across region (right) only a subset of trials with a maximum corridor distance of three are included in the analysis reported in this table.

		Pointing within region			Pointing across region		
Prediction		$t(17)$	p	d	$t(17)$	p	d
Local	First perspective in corridor	2.045	.028 *	0.48	-0.580	.715	-0.14
Region	First perspective	0.218	.415	0.05	-0.191	.575	-0.05
	Most frequent perspective	2.064	.027 *	0.49	0.177	.431	0.04
	U-shape geometry	0.422	.339	0.10	0.803	.217	0.19
Global	First perspective	0.654	.261	0.15	0.423	.339	0.10
	Most frequent perspective	0.794	.219	0.19	-1.743	.950	-0.42
	U-shape geometry	0.558	.292	0.13	2.185	.022 *	0.52

Note. For pointing across region (right) only a subset of trials with a maximum corridor distance of three are included in the analysis reported in this table.

** <.05, one-sided, larger than zero, no correction for multiple comparisons.*

Discussion

In the second experiment, a new sample of participants memorized the same regionalized environment as in Experiment 1. Again, both halves of the environment were explored separately. However, this time, participants started learning from the transition point between the two regions, walking towards the dead end of either region. Thus, in contrast to Experiment 1, participants now had the chance to encode the connection of the two regions from the beginning and integrate each additional corridor into their existing knowledge. We again observed effects in favor of local corridor units and global integration. Results regarding regional memory units are less clear.

Pointing latency increased with increasing corridor distance to the target. Similar to Experiment 1, this pattern can be explained by local corridor units that are stored in memory and connected along the learning order. Accessing the location of a target is bound to successively activating each memory unit until reaching the unit that contains the target. Each transition costs time and accuracy. In line with the distance effect, the examination of orientation dependency showed that participants were significantly faster when aligned with the first perspective experienced in each corridor, compared to the seven remaining orientations. This as well speaks for local memory units.

As in Experiment 1 participants learned the connection of both regions well, even though they never experienced the entire environment at once. Pointing accuracy depended on corridor distance, but not on whether participants pointed within or across region. In Experiment 1 we found a facilitative effect on pointing latency when recalling spatial information within one's current region compared to a more time-consuming access of targets positioned in the other region. This effect disappeared in Experiment 2. Time needed for pointing within one's current region and across regional boundaries now was of comparable length. The analysis of reference frames for within region pointing revealed—besides the local corridor alignment—a fit to a regional reference frame following the most frequently experienced perspective. This, however, could simply be an artifact of the collinearity of the consulted predictions. Accounting for the facilitative effect of local corridor alignment dissolves the facilitative effect of the most frequently experienced perspective within a region and vice versa. Since the regional reference frame analysis is ambiguous and we find no region distance effect in latency we are reluctant to argue in favor of regional memory units. We cannot exclude them with absolute certainty, but if they are present they seem less strong than in Experiment 1.

Just as in Experiment 1 results suggest that no global reference frames were used for pointing to target location within one's current region. However, again evidence for global embedding was found in the reference frame analysis of across region trials when limiting target distance to maximum three corridors. Like Experiment 1 the full set of across region trials (target distance 1–7) was indecisive with regards to any reference direction tested.

In sum, our results of Experiment 2 suggest that subjects relied on local memory units and that integration into a global memory unit might have occurred. Surprisingly, changing the learning procedure between Experiment 1 (end-to-transition point) and Experiment 2 (transition-to-end point) appears to mitigate the clustering effects of regions. And this, even though there are still other cues triggering a regionalization. Like Experiment 1, regions in Experiment 2 were still dissociated by color, semantic membership of landmarks, complexity of the angle of turn and spatio-temporal learning experience. Although the effort for global embedding was presumably lowered in Experiment 2 by allowing for a continuous integration of new spatial information, specifically when starting to learn the second region at the transition point participants are already familiar with, this did not elucidate further the findings regarding a global memory unit that were found in Experiment 1. This will be elaborated upon in the General Discussion.

General Discussion

In two experiments participants memorized an environment consisting of two connected regions by active exploration of a virtual environment. Participants either learned each region starting at the dead-end, exploring the four corridors of the region until ending up at the inter-regional transition point (Experiment 1) or they learned each region starting at the transition point, walking 'outwards' to the dead-end of each region (Experiment 2). Subsequently, participants pointed from different locations within the environment to targets within the same region or the other region. The aim of this study was to identify how the memory for this clustered space might be structured. Specifically, we aimed to establish whether participants stored the environment on a local corridor level or (also) formed memory units comprising individual regions or the environment as a whole, thus, revealing hierarchical structure of spatial memory. Additionally, we addressed different cues that might set the orientation of potentially formed reference frames on the different levels of hierarchy. In sum, we found evidence for local, regional as well as global memory units across two experiments, and different cues driving reference frame orientation.

To examine the existence of different hierarchical levels we decided for a two-fold approach when analyzing participants pointing performance. First, an analysis of pointing latency with varying corridor distance to and regional belonging of the target was conducted. The speed of memory retrieval should be telling about the memory structure, as pointing latency should increase with every new memory unit that is accessed (see also Meilinger, Strickrodt, & Bülthoff, 2016). Second, we tested how well participants' performance in different body orientations fit to a number of potential spatial reference systems, either limited to local corridors or spread across regions or the whole environment. Bodily alignment with the main orientation of the formed reference frame should facilitate memory recall (e.g., McNamara, Sluzenski, & Rump, 2008; Mou, McNamara, Valiquette, & Rump, 2004). Since measured within the same spatial task (i.e., pointing to non-visible targets located beyond ones' current vista space) both approaches should draw from the same memory source, thus, jointly add to the picture of how the spatial memory for the regionalized space is structured. Whereas the first can give insights into the number and expansion of memory uni*t*(s), the second allows to make conclusions about the nature of those memory units, that is whether they possess the spatial feature of an oriented reference system.

We found clear evidence for the formation of local memory units in both experiments and both in the analysis of distance and cluster effects and the analysis of reference frames. Pointing latency increased with increasing corridor distance to the target. This pattern can be explained by the presence of local, interconnected corridor units in memory and by a time-consuming process of successively retrieving all local units that lie between a participant's current location and the target location (see also Meilinger, Strickrodt, & Bülthoff, 2016; Pantelides, Kelly, & Avraamides, 2016). We also found quicker pointing when participants were aligned with the first experienced orientation within each local corridor, compared to being aligned with the remaining body orientations, which indicates that local reference frames were employed, one for each corridor. This was found irrespective of the learning direction (Experiment 1 and 2). Other studies observed evidence for local corridor units as well, in the form of distance effects (Meilinger et al., 2016), local alignment effects (Meilinger et al., 2014; Werner & Schmidt, 1999), and confusion errors based on vista space information (Marchette et al., 2017). The importance of local, bounded places as a core unit for the representation of large-scale space is further substantiated by our results.

Experiment 1 indicates that participants formed regional memory units as well. In addition to the local corridor effects, pointing latency also increased when pointing to targets across regions compared to pointing the same distance within a region. Accordingly, the analysis of reference frames revealed evidence for distinct regional reference frames that go beyond the facilitative effect of local corridor alignment. We found a fit to predictions of regional reference frames that follow the first experienced perspective (first corridor of region) and the salient geometry of a U-shape when concentrating on within region pointing trials. Thus, we established clustering into regional units. So far, literature concerned with hierarchical spatial representation and the formation of regions is inconsistent in its understanding of how exactly regions are represented. The format of representing regions or clusters is argued to be either rather non-spatial, such as conceptual or semantic labels (Hirtle & Jonides, 1985; Hirtle & Mascolo, 1986), as well as spatial, in the form of a topological understanding of connectivity and containment (Stevens & Coupe, 1978; Wang & Brockmole, 2003a, 2003b; Wiener & Mallot, 2003) or in the form of a metrical relational representation (Greenauer & Waller, 2010; McNamara et al., 2008; Meilinger & Vosgerau, 2010). Our study is the first (at least to our knowledge) to provide experimental particulars permitting the conclusion that two distinct reference frames accumulating multiple vista spaces can be formed in regionalized space. Jointly considering both analyses and in particular detecting regional memory units not only by latency increase beyond regional borders, but also by facilitative effects when aligned with regional reference frame levels highlights the spatial character of the regional memory unit. Thus, our results support the concept of regional memory units that are stored in the form of metric relational schemes. For example, multiple vista spaces are linked to one superordinate regional reference frame direction (McNamara et al., 2008; Meilinger & Vosgerau, 2010). Hereby, they extend results by Greenauer and Waller (2010) of micro- and macroreference frames formed for object arrays within a single vista space onto navigable space.

Experiment 2 was less conclusive about the structure of spatial memory on the regional level. In contrast to Experiment 1, the fact that the reference frame alignment effect with the most frequently experienced perspective vanishes when controlling for local corridor alignment does not allow for the conclusion that regional memory units were formed. This is consistent with the absence of a region effect on pointing latency. Corridor distance effects remain across both experiments, supporting the finding of local reference frames. Thus, following the most conservative approach we reckon that the most solid interpretation is that no or only very weak regional units have been formed in Experiment 2. This contrasts studies that succeeded in clustering a navigable space into regions by far less regional cues than those used in Experiment 2 of the current study. Utilizing semantic membership of landmarks and color alone was sufficient to affect subsequent route decisions (Schick et al., 2015; Wiener & Mallot, 2003). One tentative explanation for this inconsistency across studies might be that indeed different formats of regional memory co-occur—semantic (Hirtle & Jonides, 1985; Hirtle & Mascolo, 1986), topological (Stevens & Coupe, 1978; Wang & Brockmole, 2003a, 2003b; Wiener & Mallot, 2003), and metrical (Greenauer & Waller, 2010; McNamara et al., 2008; Meilinger & Vosgerau, 2010)—, but do not necessarily synchronize. Depending on the used task (i.e., pointing, route planning, etc.) different formats could be targeted. Building on this rationale it seems plausible to assume that the region effects found in Experiment 1 do not originate from effects of categorical belonging to semantic groups (animals vs. tools) or color (blue vs. red). In this case we would have expected to find similar effects on latency when pointing across regional boundaries in Experiment 2 as well. Nevertheless, it is beyond the scope of the current study to try and clarify this discrepancy.

Importantly, all we changed between Experiment 1 and 2 was the learning direction, and thus, participants awareness of how both regions are connected to each other at the very start of learning both regions. Although our data does not allow for strong claims, one possible explanation for the absence of regional memory units in Experiment 2 could be the opportunity to immediately relate each newly explored corridor to the existing memory structure of the first region when starting to learn the second region at the transition point again. In contrast, in Experiment 1 (end-to-transition) the immediate connection of unfamiliar corridors with existing memory structures was not possible. A novel, independent spatial unit might have been formed at the start of the second region with the new, following corridors added immediately. Restructuring the already existing two separate regions when reaching the familiar transition point into one common unit might be associated with higher mental effort compared to simply learn how the two regions are connected to each other, leading to the observed region effects. Indeed, comparable results to our study were found by Han and Becker (2014), who had participants learn landmarks located in two connected regions while varying the point of time the connecting route was introduced (e.g., immediately/very early in the learning phase vs. only after a few blocks of learning regions separately and being tested throughout) and the extend and means of encounter with it (e.g., watching a video of the connecting route vs. seeing the connecting route and parts of the other region constantly vs. actively navigating across regions). Slower pointing latencies for across region pointing where only found when the connecting route was introduced in later blocks, while enabling immediate and very early encounter always led to comparable pointing latencies for within and across region pointing. From studies investigating film and narrative comprehension it is known that, among others, (unexplained) changes in spatial locations of the protagonist negatively effects reading time of a narrative (e.g., Rinck & Weber, 2003; Scott Rich & Taylor, 2000), film cuts alter comprehension and memory of the movie (Schwan, Garsoffky, & Hesse, 2000), and temporal shifts in a narrative can weaken memory binding of pre- and post-shift content (Ezzyat & Davachi, 2011; Zwaan, 1996; Zwaan, Langston, & Graesser, 1995). Hence, strong discontinuities in an episode alter how such content is perceived and remembered. This is in line with, for example some associative memory models that suggest that (un)available shared context affects associative binding between incoming information (Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009a, 2009b). Translated to the context of this study, not having access to an immediate reference to the previously learned region when stating to learn the second region in Experiment 1 may both lead to the detection of a spatial discontinuity ("I am somewhere else") as well as stronger temporal discontinuity compared to Experiment 2 because the transition point is discovered later in time. In sum, the availability of a reference point early on during learning connected spaces that at best allows for a continuous flow of incoming information across regional boundaries might be a very crucial factor for the integration of the two spaces and, vice versa, the unavailability thereof determine clustering in memory. Although this explanation fits to the pattern observed across the two experiments, further research is necessary to understand the processes of regional clustering. For example, alternatively the walking direction itself, namely whether regions were first explored on a convergent (end-to-transition point, Experiment 1) or on a divergent path (transition-to-end, Experiment 2), irrespective of the time the transition point was learned, could alter memory formation.

In both Experiments we observed evidence for the integration of all eight corridors into a global memory unit. In Experiment 1 global reference frames following the first experienced perspective and the U-shape of the first learned region were indicated by the reference frame analysis of across region trials, in Experiment 2 for the U-shape. However, this was only the case when reverting to trials querying corridor distance one to three, leaving out trials with higher corridor distances. Individual standard deviation suggests that this might be due to the increased noise in the across region data of all distances compared to within region trials were only corridor distance one, two and three were tested. It should be noted here that the analysis of distance and cluster effects is tailored to detect either local and regional memory units or, in contrast, the absence of local and regional memory units and the exclusive presence of one global memory unit. In the latter case, no increase in latency across corridors or the regional boundary would have been expected, since all spatial information was accessible with similar ease from within one common memory unit as when learning from a city map (Frankenstein, Mohler, Bülthoff, & Meilinger, 2012). However, we observed distance and regional cluster effects that are in favor of local and regional memory units; and based on these results alone no conclusions about the formation of an additional global memory unit can be made. Therefore, consulting results from the reference frame analysis is essential. Interestingly, jointly considering results from the distance and cluster analysis and from the reference frame alignment for across region pointing covering distance one to three reveals successive activation of local memory units and the formation of a global reference frame. This indicates that although a global reference frame is used for recalling across region information no simple all-at-once readout from this memory unit occurred. Rather, even a retrieval process that is based on a global embedding seem to be bound to successive activation of local memory units. A similar interpretation can be made based on the observation of local and regional memory units for within region pointing in Experiment 1. Having access to and using a regional reference frame does still involve successive activation of local memory units, leading to both local and regional effects. If successive activation of local memory units is essential even in the presence of and during utilization of higher order memory units it, first of all, raises the question of what exactly is stored in a global (and regional) reference frame if this information cannot be used for an easy and fast read-out. Secondly, it might explain why including higher corridor distances led to non-conclusive results for across region trials. The successive integration of local memory units—even though stored additionally within a global reference system—at the moment of testing might be subject to limitations of working memory capacity. Utilizing both representations from local and global memory units might functions exclusively as long as this can be done within working memory capacity. The process, however, cannot be upheld when going beyond the limit at a specific corridor distance, therefore, our effects vanish with corridor distances higher than three. Considering Figure 2 and 5 this might well be the case at around corridor distance 4 or 5, where the distance effect on latency seems to flatten out. It remains a question for further investigation what exactly is happening to the spatial recall process if working memory limit is reached.

We cannot completely rule out an alternative explanation for the fit to a global reference frame prediction for smaller distances in across region pointing, namely, that no global but an additional memory unit on the regional level was formed which is not covering the whole environment, but only the corridors around the transition point. Indeed, across region trials covering maximum corridor distance of three are limited to the area around the transition point, namely from there three corridors of the blue and three of the red region, leaving out the outermost corridors. However, if another unit on the regional level was formed we would expect to find a fit to local reference frame predictions as well, just as for pointing within region. As this was not the case, formation of a global reference frame seems more likely.

The use of global reference frames for pointing across regional boundaries only is in line with findings on spatial layouts learned in vista space by Greenauer and Waller (2010) and extend them to environmental space. Greenauer and Waller (2010) found evidence for a macroreference frame when pointing across two separate object layouts that were learned in a single room, but not when pointing within one object array. Particularly, the study implies that superordinate reference frames might only be used when pointing to a target which is located beyond the scope of a smaller microreference frame. Likewise, in our study global reference frames seem to be accessed flexibly only when required, namely, when recalling the relative direction of a target in the other region.

As pointed out before, the reported effects indicating local and regional reference frames are restricted to trials of within region pointing, while pointing trials that target objects within the other region show recall behavior independent of local and regional main orientations. Importantly, even though we also see corridor distance effects in across region trials, here no evidence for local reference frames can be detected. In other words, even though the distance effect supports the use of local memory units when pointing across regional boundaries, the facilitative effect of local alignment disappears as soon as participant point to a target within the other region. An attempted explanation could be that, when pointing across region, the corridor orientation visible from one's current location might conflict with the obliquely oriented geometry of the other region, if visualized in the same reference frame. A study by Meilinger and Bülthoff (2013) suggests that visual pointing can lead to interference between the surrounding visual geometry and the geometry of the memorized environment. However, this interpretation is speculative. Whether and how visual input can interfere with selected reference frames, however, cannot be resolved based on our results.

What is setting the direction of a regional reference frame? Former studies concerned with the formation of a reference frame covering multiple vista spaces focused on the importance of the first perspective taken in the navigable space that must be memorized (e.g., Richardson, Montello, & Hegarty, 1999; Tlauka, Carter, Mahlberg, & Wilson, 2011; Wilson, Wilson, Griffiths, & Fox, 2007). In these studies, the first corridor walked was typically also the longest and/or parallel to the last corridor, forming a salient geometric U-shape. Thus, the salience of the first perspective was partially associated with the most frequently exposed perspective and the parallel environmental structure. In Experiment 1 we found support for the influence of the first perspective as well as the U-shape for setting the regional and global reference frame direction and Experiment 2 replicated the fit to a global reference frame following the U-shape of the first learned region. This indicates that the representation of navigable space might not only be shaped by initial views (first corridor) that form the basis of a reference frame into which subsequent spatial information (following corridors) are integrated into (assimilation), but by spatial information gathered across a sequence of corridors which, in combination, can form a new distinct reference direction (accommodation) (Piaget & Inhelder, 1969). This is in line with an outlook given by McNamara and Valiquette (2004) after outlining their theoretical framework of spatial reference systems, where they state “the first segment is a strong candidate because of the salience conferred by novelty”, but furthermore “It is also possible that one or more of the other segments of the path might be used to establish a reference system” (pp. 21-22). Many studies examined the factors influencing the alignment of the mental reference frame in vista space (e.g., Kelly & McNamara, 2008; Mou & McNamara, 2002; Shelton & McNamara, 2001; Valiquette & McNamara, 2007). Our study highlights, that comparable efforts should now be taken to shed light on the factors influencing reference frame selection in navigable space as well.

We realize that one must remain cautious in making too strong interpretation based on the analysis of reference frames alone, and in particular, about the evidence for regional and global reference frames. Significant fits reported are based on one-sided t-tests, without correction for multiple tests evaluating the three theoretical predictions (first perspective, most frequent, U-shape) on regional and global level for within and across pointing. Nevertheless, both the separation of within and across trials and the three potential cues for reference frame alignment are theoretically motivated and specified as directional hypotheses. Furthermore, results of local and global reference frames could be replicated across two experiments. And most importantly, the reference frame results are in high accordance with the analysis of distance and region effects.

Participants were well able to learn the environment. Pointing accuracy did not depend on whether participants pointed to targets within one's current region or to targets within the other region. And this, despite separate learning of both regions and despite a single, comparatively complex (45°) common reference point at the transition between the two halves of the environment. Previous studies already showed the rapid learning of spatial relations between two areas immediately after being exposed to a connecting route (e.g., Ishikawa & Montello, 2006; Schinazi, Nardi, Newcombe, Shipley, & Epstein, 2013) and our study is in line with those results. In addition, we present evidence that global metric embedding can take place without ever walking the entire environment at once.

Substantiating the presence of local, regional and global reference frames, our study promotes hierarchical concepts of spatial memory (e.g., Mallot & Basten, 2009; Stevens & Coupe, 1978; Wiener & Mallot, 2003). It should not be left unsaid that the current analysis is indecisive of whether multiple hierarchical levels occur within a single participant or only across participants. The analysis is based on the average pointing performance of the whole sample. Efforts to identify single level or hierarchical memory structures within single participants failed to produce interpretable results. Also, it was not possible to separate participants in local-only, region-only or global-only groups based on their performance. It remains an issue of further examination to ascertain whether multiple reference frames are used by single individuals (i.e., hierarchical representation), or whether our results mirror the average of individual strategies across participants, where each participant relies exclusively on the local, regional or global level.

In past research going beyond the immediately visible surrounding of vista space uncovered multiple aspects of spatial learning. Opaque borders seem to distort distance judgements (e.g., Kosslyn et al., 1974; Newcombe & Liben, 1982), affect online updating of landmark locations (e.g., Avraamides & Kelly, 2010; Wang & Brockmole, 2003, 2003) and lead to latency costs when switching between spatial units (e.g., Brockmole & Wang, 2002, 2003) or when pointing to targets located in increasingly distant corridors (Meilinger et al., 2016). Also, exact spatial knowledge of a target within a vista space does not coincide with a similarly good knowledge about where this vista space itself is located (Marchette et al., 2017). Additionally, individual reference frames were found to be formed for individual vista spaces (Meilinger et al., 2014; Werner & Schmidt, 1999). Considered jointly these findings suggest that visual boundaries enclosing vista spaces seem to serve as molds for local spatial memory units. Results of these studies can be well explained by non-hierarchical theories postulating the formation of local memory units and their successive connection (e.g., Chrastil & Warren, 2014; Meilinger, 2008; Warren, Rothman, Schnapp, & Ericson, 2017). Non-hierarchical theories assuming not multiple local, but a single global representation of all spatial information cannot account for these results.

Notwithstanding, there is a large number of studies indicating hierarchical structures in human spatial memory. For example, Stevens and Coupe (1978) showed that the relative position between remote cities seemed to be judged based on the relative position of the federal states the cities are in rather than their actual location. Both for vista (McNamara, Hardy, & Hirtle, 1989) and environmental space (Hirtle & Jonides, 1985) hierarchical clusters were unveiled based on participants landmark recall protocols. Landmark proximity in these hierarchical clusters were associated with distorted distance estimations between landmarks (Hirtle & Jonides, 1985; McNamara et al., 1989) and priming effects in a recognition task (McNamara et al., 1989). Evidencing clustering of spatial information is a prerequisite for hierarchical organization of space. However, a common concern regarding these studies is, whether the results could likewise be explained by non-hierarchical theories when assuming that object locations are indeed clustered, but they are not organized in a hierarchical fashion. In this case, judgements of relation and distance as well as priming effects might only reflect a distorted memory misrepresenting physical space, but not necessarily a memory that does possess another level of hierarchy for regions and clusters.

We suspect that to make profound interpretations regarding hierarchical spatial representations often additional dependent variables and/or approaches need to be consulted, as was done in a number of other studies. For example, McNamara (1986) contrasted participants distance judgements between landmarks (reflecting possible distortions in memory) and spatial priming effects in a recognition task after participants learned a clustered vista space containing four regions. While non-hierarchical theories would predict an exponential decay of priming effects with psychological distances this was not reflected in the data. Thus, priming effects seemed to not simply mirror an erroneous representation distorted according to formed clusters, but a hierarchical representation. In a route choice task Wiener and Mallot (2003) showed that participants tended to approach the region containing the target object directly rather than choosing a path with a longer dwelling time in the non-target region but equivalent in length and complexity. Importantly, for one of the environments they used it seems reasonable to assume that the two alternative routes would also lead to equally long psychological route distances. The environment was a grid field of two rectangles arranged opposite each other that formed two regions. One would expect a similar distortion in spatial memory within both regions, potentially towards the centroids of each region (e.g., Huttenlocher, Hedges, & Duncan, 1991), thus, distortions in memory could not account for a bias in route selection. Still detecting region effects on route planning suggests an additional memory layer for representing regions. Our results line up nicely with these studies by highlighting local, regional and global effects with two complementary analyses that jointly indicate hierarchical structures in human spatial memory. Latency increase across corridor distance and regional boundaries indicate that new memory units must have been activated. The reference frame analysis supports the effects further and acknowledges the spatial character of the units, while complementing them with evidence for global reference frames. We reckon that non-hierarchical theories have difficulties explaining these results.

Following an alternative approach, the superordinate level, which manifests in regional boundary and regional and global alignment effects, might not consist of a coordinate system that is yielding metric embedding of subordinate units within a region or across the environment, but rather of a common reference direction or vector. This does not touch the assumption that memory units for vista spaces are stored in the form of locally confined, spatial reference frames, leading to the observed distance and local alignment effects. However, the encountered regional and global effects could also be caused by additional regional or one additional global reference vector stored in memory on a single superordinate level, a main orientation extrapolated across a limited or the entire number of vista spaces. Like the spatial reference system proposed by Shelton and McNamara (2001) the direction itself can be described as a conceptual "north", a privileged direction in the environment. It might be set, for example, by the first segment walked (first perspective alignment), or by salient inner-regional structures such as a U-shape. However, in contrast to Shelton and McNamara (2001), the superordinate vector is different in a sense that spatial relations are not explicitly specified with respect to a spatial reference system as coordinates, but rather it reflects an anchor orientation propagated across multiple corridors, for example, via a global sense of direction system (Sholl, Kenny, & DellaPorta, 2006). Bodily alignment with this superordinate vector does allow for an easy access of the remaining object locations not due to the availability of and alignment with a spatial reference system that contains the relative position of objects located beyond one's current vista space, but rather because the superordinate vector facilitates the coordination and alignment of the local memory units stored on the subordinate level. Such a vector approach could explain the fact that distance effects prevail—suggesting successive activation of local memory unit, rather than an all-at-once readout—in the presence of regional and global orientation dependencies, as found in our study. However, based on a global sense of direction system (Sholl et al., 2006) it also limits the hierarchical representation to a maximum of two levels, a local level and an additional vector. This would indeed imply that the detection of regional and global memory units in Experiment 1 are due to individual differences, meaning that some participants formed two separate regional vectors, others formed a single global vector. This would be an additional conjecture to explain our results. At the same time this approach is more economic than assuming an additional third level to be represented as in the case of the hierarchical representation of local, regional and global reference frames. Such a superordinate vector could account for alignment effects found across multiple vista spaces and still allow for globally inconsistent spatial memory and biases (e.g., Warren, Rothman, Schnapp, & Ericson, 2017) as only orientation, but not location is specified on the superordinate level. Forming such a vector across multiple corridors during the learning phase should involve accumulation of error during updating, thus, accuracy of this vector should decrease with increasing corridor distance. Besides the aspect stated earlier, namely, that limited working memory capacity affects successive recall of local units, also error accumulation during learning could explain why higher corridor distances for across region pointing led to inconclusive results about a uniform global reference direction. In that sense a superordinate spatial reference system might consist of a full-blown coordinate system or a single reference direction. Although such an approach might be uncommon and in need of additional theoretical polishing, we regard it worth further investigation in future experiments.

Taken together, we find that participants encoded navigation spaces on the local (Experiment 1 and 2), regional (Experiment 1) and global level (Experiment 1 and 2). The presence of regional reference frames indicated in the alignment effects suggest that multiple vista spaces that form clearly circumscribed areas may serve as molds for spatial memory units on the regional level, and that the representation of regions is not limited to conceptual or topological knowledge (Hirtle & Mascolo, 1986; Meilinger, 2008; Wang & Brockmole, 2003a, 2003b; Wiener & Mallot, 2003), but can indeed rely on distinct mental reference systems subsuming multiple places, thus, possessing a spatial character (e.g., McNamara et al., 2008; Meilinger & Vosgerau, 2010). The segmentation into distinct spatial reference systems on the regional level might be dependent on the learning procedure. Experiment 2, which allowed for an immediate relation of new spatial information to existing knowledge by starting to learn the second region at a familiar reference point, does not seem to trigger the formation of regionally confined spatial reference systems as strongly as being introduced to a common reference point between two areas at a later time during learning (Experiment 1). Thus, the point in time when the connectivity between separately learned spaces is introduced might play a role for the emerging structure of spatial memory. Spatial information of the entire environment was found to be integrated on a global level. The use of this level, however, was dependent on the relationship between one's current position and the target location (i.e., only for across region pointing), indicating that the stored mental structures representing navigable space are not consulted exhaustively every time spatial information are recalled but used in customized fashion instead. Importantly, our findings demonstrate that the representation of navigable space is not limited to local memory units encompassing single corridors or streets (e.g., Chrastil & Warren, 2014; Meilinger, 2008; Warren, Rothman, Schnapp, & Ericson, 2017). Similarly, the evidence for local and regional memory units besides support for a global embedding speaks against a purely global integration into a single, common reference frame (e.g., Gallistel, 1990; Ishikawa & Montello, 2006; O'Keefe & Nadel, 1978; Sholl, 2001).

Ultimately, our results show that object locations in navigable space seem to be represented flexibly on different levels (maybe not necessarily within a single individual) and thereby support hierarchical theories of spatial memory (e.g., Mallot & Basten, 2009; Stevens & Coupe, 1978; Wiener & Mallot, 2003). We found strong support that local memory units seem to be a key component for memorizing navigable space, as they are pervasive across both experiments. Potentially, they form the basic units in spatial memory, that subsequently can be consolidated to form memory units on one or more superordinate levels, for example, in the form of regional and global reference frames.

The processes involved in the generation of such hierarchical representation and how the memory units interact within and across hierarchical levels will have to be clarified in future research.

Acknowledgments

We would like to thank Joachim Tesch for help with the virtual environment setup.

Supplementary Material

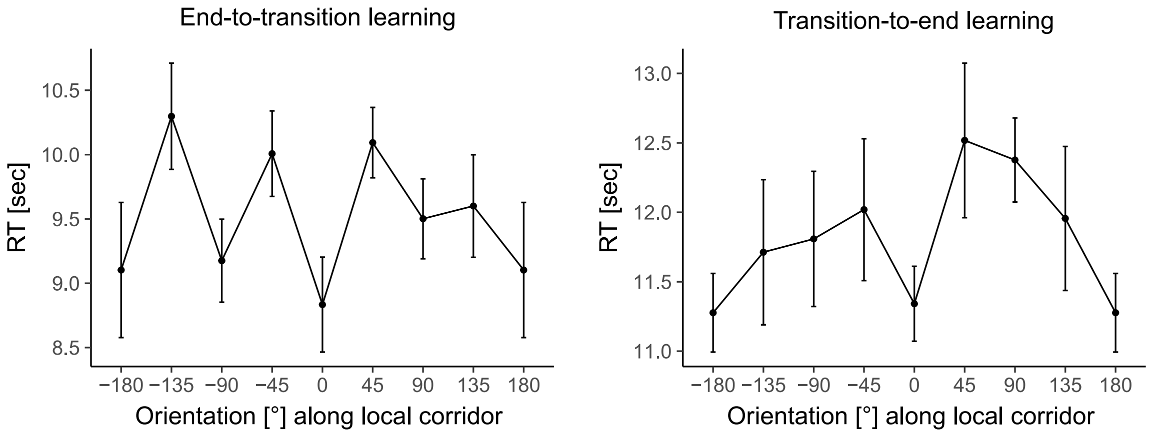


Figure S1. An overview of the pattern of pointing latency with respect to each local corridor. Pointing latency as a function of eight possible body orientations during execution of pointing, for Experiment 1 (end-to-transition, left) and for Exeriment 2 (transition-to-end, right). Orientation 0° corresponds to the first perspective experienced within each corridor when exploring the environment, and should lead to fastest pointing performance compared to the remaining orientations in accordance with the prediction of local reference frames (compare to Figure 3, predictions). Deviations from this orientations were made in steps of 45° either clockwise (+) or counterclockwise (-). ±180° is depicted twice. Means and SEMs are depicted. Only data points from within region pointing are included, corrected for effect of distance. Mean reaction time of 9.85 sec for Experiment 1 and 12.05 sec for Experiment 2 was added to the residuals to make them easier to interpret.

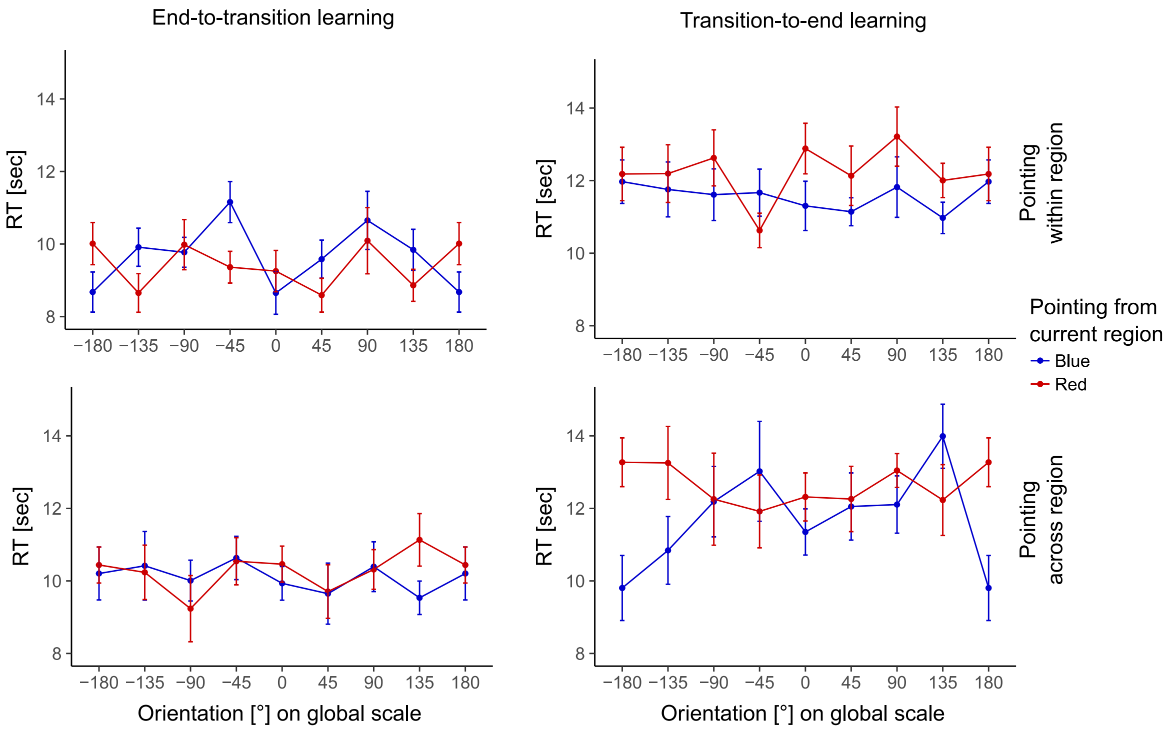


Figure S2. Visualization of the pattern of pointing latency with respect to the whole environment. Pointing latency as a function of orientation on global scale, for Experiment 1 (end-to-transition, left) and for Exeriment 2 (transition-to-end, right). Orientations 0°, ±90° and 180° correspond to the four main orientations of the blue region, orientations ±45° and ±135° correspond to the main orientations of the red region (compare to compass rose shown in Figure 3, predictions). ±180° is depicted twice. Means and SEMs are shown. Only data points from within region pointing are included, corrected for effect of distance. Mean reaction time of 9.85 sec for Experiment 1 and 12.05 sec for Experiment 2 was added to the residuals to make them easier to interpret. The figure allows to observe the latency pattern for each region separately. In general, reverse patterns for the obliquely aligned regions with best pointing performances along 0°, ±90° and 180° for the blue region and along ±45° and ±135° for the red region would be in favor for local/regional reference frames. An approximation of the patterns of both regions would be in favor for a global embedding. Note that this figure was included to give an idea of the latency pattern. For a meaningful interpretation we concentrated on clear predictions that we tested for in the result sections of the manuscript.

References

Avraamides, M. N., & Kelly, J. W. (2010). Multiple systems of spatialmemory: Evidence from described scenes. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 635–645. http://dx.doi.org/10.1037/a0017040

Brockmole, J. R., & Wang, R. F. (2002). Switching between environmental representations in memory. Cognition, 83, 295–316. http://dx.doi.org/10.1016/S0010-0277(02)00012-4

Brockmole, J. R., & Wang, R. F. (2003). Changing perspective within and across environments. Brain & Development, 25, 291–293.

Byrne, R. W. (1979). Memory for urban geography. The Quarterly Journal of Experimental Psychology, 31, 147–154. http://dx.doi.org/10.1080/14640747908400714

Chrastil, E. R., & Warren, W. H. (2014). From cognitive maps to cognitive graphs. PLoS ONE, 9, e112544. http://dx.doi.org/10.1371/journal.pone.0112544

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. Science, 311, 381–384. http://dx.doi.org/10.1126/science.1121739

Dillon, M. R., Huang, Y., & Spelke, E. S. (2013). Core foundations of abstract geometry. Proceedings of the National Academy of Sciences of the United States of America, 110, 14191–14195. http://dx.doi.org/10.1073/pnas.1312640110

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? Psychological Science, 22, 243–252. http://dx.doi.org/10.1177/0956797610393742

Frankenstein, J., Mohler, B. J., Bülthoff, H. H., & Meilinger, T. (2012). Is the map in our head oriented north? Psychological Science, 23, 120–125.

Gallistel, C. R. (1990). The organization of learning. Cambridge, MA: MIT Press.

Greenauer, N., & Waller, D. (2010). Micro- and macroreference frames: Specifying the relations between spatial categories in memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 938–957. http://dx.doi.org/10.1037/a0019647

Han, X., & Becker, S. (2014). One spatial map or many? Spatial coding of connected environments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 40, 511–531. http://dx.doi.org/10.1037/a0035259

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. Nature, 370, 57–59. http://dx.doi.org/10.1038/370057a0

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. Cognition, 61, 195–232. http://dx.doi.org/10.1016/S0010-0277(96)00714-7

Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. Memory & Cognition, 13, 208–217. http://dx.doi.org/10.3758/BF03197683

Hirtle, S. C., & Mascolo, M. F. (1986). Effect of semantic clustering on the memory of spatial locations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 182–189. http://dx.doi.org/10.1037/0278-7393.12.2.182

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. Journal of Mathematical Psychology, 46, 269–299. http://dx.doi.org/10.1006/jmps.2001.1388

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. Psychological Review, 98, 352–376. http://dx.doi.org/10.1037/0033-295X.98.3.352

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. http://dx.doi.org/10.1016/j.cogpsych.2005.08.003

Jeffreys, H. (1961). Theory of probability (3rd ed.). Oxford, UK: Oxford University Press.

Kelly, J. W., & McNamara, T. P. (2008). Spatial memories of virtual environments: How egocentric experience, intrinsic structure, and extrinsic structure interact. Psychonomic Bulletin & Review, 15, 322–327.http://dx.doi.org/10.3758/PBR.15.2.322

Kelly, J. W., & McNamara, T. P. (2010). Reference frames during the acquisition and development of spatial memories. Cognition, 116, 409–420. http://dx.doi.org/10.1016/j.cognition.2010.06.002

Kosslyn, S. M., Pick, H. L., Jr., & Fariello, G. R. (1974). Cognitive maps in children and men. Child Development, 45, 707–716. http://dx.doi.org/10.2307/1127837

Lee, M. D., & Wagenmakers, E.-J. (2013). Bayesian cognitive modeling: A practical course. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139087759

Mallot, H. A., & Basten, K. (2009). Embodied spatial cognition: Biological and artificial systems. Image and Vision Computing, 27, 1658–1670. http://dx.doi.org/10.1016/j.imavis.2008.09.001

Marchette, S. A., Ryan, J., & Epstein, R. A. (2017). Schematic representations of local environmental space guide goal-directed navigation. Cognition, 158, 68–80. http://dx.doi.org/10.1016/j.cognition.2016.10.005

McNamara, T. P. (1986). Mental representations of spatial relations. Cognitive Psychology, 18, 87–121. http://dx.doi.org/10.1016/0010-0285(86)90016-2

McNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 211–227. http://dx.doi.org/10.1037/0278-7393.15.2.211

McNamara, T. P., Sluzenski, J., & Rump, B. (2008). Human spatial memory and navigation. In H. L. Roediger, III (Ed.), Learning and memory: A comprehensive reference (Vol. 2, pp. 157–178). Oxford, UK: Elsevier. http://dx.doi.org/10.1016/B978-012370509-9.00176-5

McNamara, T. P., & Valiquette, C. M. (2004). Remembering where things are. In G. L. Allen (Ed.), Human spatial memory: Remembering where (pp. 3–24). Mahwah, NJ: Erlbaum.

Meilinger, T. (2008). The network of reference frames theory: A synthesis of graphs and cognitive maps. In C. Freksa, N. S. Newcombe, P. Gärdenfors, & S. Wölfl (Eds.), Spatial cognition VI. Learning, reasoning, and talking about space (Vol. 5248 LNAI, pp. 344–360). Berlin, Germany: Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-87601-4_25

Meilinger, T., & Bülthoff, H. H. (2013). Verbal shadowing and visual interference in spatial memory. PLoS ONE, 8, e74177. http://dx.doi.org/10.1371/journal.pone.0074177

Meilinger, T., Frankenstein, J., Watanabe, K., Bülthoff, H. H., & Hölscher, C. (2015). Reference frames in learning from maps and navigation. Psychological Research, 79, 1000 –1008. http://dx.doi.org/10.1007/s00426-014-0629-6

Meilinger, T., Riecke, B. E., & Bülthoff, H. H. (2014). Local and global reference frames for environmental spaces. Quarterly Journal of Experimental Psychology, 67, 542–569. http://dx.doi.org/10.1080/17470218.2013.821145

Meilinger, T., Strickrodt, M., & Bülthoff, H. H. (2016). Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation. Cognition, 155, 77–95. http://dx.doi.org/10.1016/j.cognition.2016.06.003

Meilinger, T., & Vosgerau, G. (2010). Putting egocentric and allocentric into perspective. In C. Hölscher, T. F. Shipley, M. Olivetti Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), Spatial cognition VII, spatial cognition 2010: lecture notes in computer science (Vol. 6222 LNAI, pp. 207–221). Berlin, Heidelberg: Springer.

Moar, I., & Bower, G. H. (1983). Inconsistency in spatial knowledge. Memory & Cognition, 11, 107–113. http://dx.doi.org/10.3758/BF03213464

Montello, D. R. (1993). Scale and multiple psychologies of space. In A. U. Frank & I. Campari (Eds.), Spatial information theory: A theoretical basis for GIS (Vol. 716, pp. 312–321). Berlin, Heidelberg: Springer.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 162–170. http://dx.doi.org/10.1037/0278-7393.28.1.162

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 142–157. http://dx.doi.org/10.1037/0278-7393.30.1.142

Newcombe, N., & Liben, L. S. (1982). Barrier effects in the cognitive maps of children and adults. Journal of Experimental Child Psychology, 34, 46–58. http://dx.doi.org/10.1016/0022-0965(82)90030-3

O'Keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. Oxford, UK: Oxford University Press.

Pantelides, S. N., Kelly, J. W., & Avraamides, M. N. (2016). Integration of spatial information across vision and language. Journal of Cognitive Psychology, 28, 171–185. http://dx.doi.org/10.1080/20445911.2015.1102144

Piaget, J., & Inhelder, B. (1969). The psychology of the child. New York, NY: Basic Books.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009a). A context maintenance and retrieval model of organizational processes in free recall. Psychological Review, 116, 129–156. http://dx.doi.org/10.1037/a0014420

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009b). Task context and organization in free recall. Neuropsychologia, 47, 2158–2163. http://dx.doi.org/10.1016/j.neuropsychologia.2009.02.013

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. Journal of Experimental Psychology: General, 134, 73–92. http://dx.doi.org/10.1037/0096-3445.134.1.73

Rich, S. S., & Taylor, H. A. (2000). Not all narrative shifts function equally. Memory & Cognition, 28, 1257–1266. http://dx.doi.org/10.3758/BF03211825

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. Memory & Cognition, 27, 741–750. http://dx.doi.org/10.3758/BF03211566

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 1157–1165. http://dx.doi.org/10.1037/0278-7393.15.6.1157

Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. Memory & Cognition, 31, 1284–1292. http://dx.doi.org/10.3758/BF03195811

Schick, W., Halfmann, M., & Mallot, H. A. (2015). How to construct a linguistic landmark: Language cues in the formation of hierarchical representations of space. Cognitive Processing, 16, 383–388. http://dx.doi.org/10.1007/s10339-015-0722-9

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, 515–528. http://dx.doi.org/10.1002/hipo.22111

Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? Memory & Cognition, 28, 214–223. http://dx.doi.org/10.3758/BF03213801

Shelton, A. L., & McNamara, T. P. (2001). Systems of spatial reference in human memory. Cognitive Psychology, 43, 274–310. http://dx.doi.org/10.1006/cogp.2001.0758

Sholl, M. J. (2001). The role of a self-reference system in spatial navigation. In D. R. Montello (Ed.), Cosit 2001 (pp. 217–232). Berlin, Germany: Springer Berlin Heidelberg.

Sholl, M. J., Kenny, R. J., & DellaPorta, K. A. (2006). Allocentric-heading recall and its relation to self-reported sense-of-direction. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 516–33. http://dx.doi.org/10.1037/0278-7393.32.3.516

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. American Scientist, 57, 421–457. http://www.ncbi.nlm.nih.gov/pubmed/5360276

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. Cognitive Psychology, 10, 422–437. http://dx.doi.org/10.1016/0010-0285(78)90006-3

Tlauka, M., Carter, P., Mahlberg, T., & Wilson, P. N. (2011). The first perspective alignment effect: The role of environmental complexity and familiarity with surroundings. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 64, 2236–2250. http://dx.doi.org/10.1080/17470218.2011.586710

Tversky, B. (1981). Distortions in memory for maps. Cognitive Psychology, 13, 407–433. http://dx.doi.org/10.1016/0010-0285(81)90016-5

Valiquette, C., & McNamara, T. P. (2007). Different mental representations for place recognition and goal localization. Psychonomic Bulletin & Review, 14, 676–680. http://dx.doi.org/10.3758/BF03196820

Wang, R. F. (2016). Building a cognitive map by assembling multiple path integration systems. Psychonomic Bulletin & Review, 23, 692–702.http://dx.doi.org/10.3758/s13423-015-0952-y

Wang, R. F., & Brockmole, J. R. (2003a). Human navigation in nested environments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 398–404. http://dx.doi.org/10.1037/0278-7393.29.3.398

Wang, R. F., & Brockmole, J. R. (2003b). Simultaneous spatial updating in nested environments. Psychonomic Bulletin & Review, 10, 981–986.http://dx.doi.org/10.3758/BF03196562

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. http://dx.doi.org/10.1016/j.cognition.2017.05.020

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, 669–682. http://dx.doi.org/10.1037/a0035261

Werner, S., & Schmidt, K. (1999). Environmental reference systems for large-scale spaces. Spatial Cognition and Computation, 1, 447–473. http://dx.doi.org/10.1023/A:1010095831166

Wertheimer, M. (1923). Laws of organization in perceptual forms. A source book of Gestalt psychology. Retrieved from http://psychclassics.yorku.ca/Wertheimer/Forms/forms.htm

Wiener, J. M., & Mallot, H. A. (2003). "Fine-to-coarse" route planning and navigation in regionalized environments. Spatial Cognition and Computation, 3, 331–358. http://dx.doi.org/10.1207/s15427633scc0304_5

Wilson, P. N., Wilson, D. A., Griffiths, L., & Fox, S. (2007). First perspective spatial alignment effects from real-world exploration. Memory & Cognition, 35, 1432–1444. http://dx.doi.org/10.3758/BF03193613

Zhang, H., Mou, W., McNamara, T. P., & Wang, L. (2014). Connecting spatial memories of two nested spaces. Journal of Experimental Psychology: Learning, Memory, and Cognition, 40, 191–202. http://dx.doi.org/10.1037/a0033776

Zwaan, R. A. (1996). Processing narrative time shifts. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1196–1207. http://dx.doi.org/10.1037/0278-7393.22.5.1196

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. Psychological Science, 6, 292–297. http://dx.doi.org/10.1111/j.1467-9280.1995.tb00513.x