Exploring the structure of spatial representations

Tamas Madl[a,b,*], Stan Franklin[c], Ke Chen[a], Robert Trappl[b], Daniela Montaldi[d]

[a]*School of Computer Science, University of Manchester, Manchester M13 9PL, UK*
[b]*Austrian Research Institute for Artificial Intelligence, Vienna A-1010, Austria*
[c]*Institute for Intelligent Systems, University of Memphis, Memphis TN 38152, USA*
[d]*School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK*

Abstract

It has been suggested that the map-like representations that support human spatial memory are fragmented into sub-maps with local reference frames, rather than being unitary and global. However, the principles underlying the proposed structure of these 'cognitive maps' are not well understood.

We propose that the structure of the representations of navigation space arises from clustering, i.e. from a process that groups together objects that are close in a given psychological space, and we present evidence for this claim based on participants' long-term spatial memories regarding buildings in real-world, as well as virtual reality, environments. We compare plausible dimensions of this psychological space, including spatial distance, visual similarity and functional similarity, and report strong correlations between these dimensions and the grouping probability in participants' spatial map structures, which empirically support the clustering hypothesis.

In addition, we also present the first formal predictive model of human navigation-scale spatial representation structure, based on the Bayesian cognition paradigm, and show that this probabilistic model of clustering, when provided with information regarding psychological spaces, learned from subjects, allows the prediction of their cognitive map structures for the first time.

Keywords:
Spatial representations, cognitive maps, hierarchical cognitive maps, spatial structure, spatial memory, computational cognitive modeling

1. Introduction

There has been considerable research on spatial representations facilitating navigation since Tolman coined the term 'cognitive map' (Tolman, 1948). Since then, the neural bases of such allocentric (world-centered) representations of space have been identified in rats (O'Keefe & Nadel, 1978; McNaughton et al., 2006) and humans (Ekstrom et al., 2003; Barry et al., 2006) and have been shown to play a vital role in representing locations within the environment in long-term memory. Instead of learning a single spatial map with a global reference frame, as proposed originally (Tolman, 1948; O'Keefe & Nadel, 1978), humans (as well as some non-human animals) seem to form structured spatial maps, consisting of multiple 'sub-maps', i.e. multiple representations containing spatial information about sub-sets of objects in the environment, with separate local frames of reference.

Behavioural evidence has suggested that human spatial maps are structured, and has been interpreted as comprising multi-level hierarchies (Hirtle & Jonides, 1985; McNamara, 1986; McNamara et al., 1989; Holding, 1994; Wiener & Mallot, 2003), or at least as having multiple local reference frames (Meilinger et al., 2014; Greenauer & Waller, 2010). These hierarchies, extracted from recall sequences, can be observed even in the case of randomly distributed objects with no boundaries (McNamara et al., 1989), with participants' response

[*] tamas.madl@gmail.com

times and accuracies being affected by this structure (subjects overestimated distances between objects on different branches of the hierarchy and underestimated distances within branches, and showed shorter response times for within-branch judgements). Further evidence for the existence of multiple representations in different spatial reference frames (Greenauer & Waller, 2010; Shelton & McNamara, 2001; Meilinger et al., 2014) has been derived from the accuracies of judgements of relative direction, which are heavily affected by subjects' frames of reference.

In addition to behavioural data, there is also strong neuroscientific evidence for hierarchical spatial representations (Brun et al., 2008; Kjelstrup et al., 2008), and for fragmentation into sub-maps (Derdikman & Moser, 2010) in mammalian brains. Finally, organized and structured maps (instead of a single representation) are consistent with 'chunking' long-term memory (Gobet et al., 2001) and with hierarchical models of cognition (Cohen, 2000), and have multiple information processing advantages, including the increased speed and efficiency of retrieval search, and economical storage.

The rate at which results about structured cognitive maps (navigation-space allocentric representations) in humans have been published has declined since the pioneering work of the eighties and nineties, partly because of some controversy surrounding the term 'cognitive map' [1]. The methodological difficulties plaguing behavioural research into the organization of cognitive maps are additional likely reasons for this decline. Unfortunately, humans do not have introspective insight into the structure of their cognitive maps. Thus, map structure can only be inferred indirectly, with a small set of possible behavioural paradigms such as those tapping recall patterns or priming effects, which are prone to noise (see Section 4, General Discussion, for a comparison of advantages and disadvantages of different methods).

Although map structure is not introspectively accessible nor immediately apparent, it does play an important role in spatial cognition. It has been shown in experiments involving priming, distance and angle estimations, and sketch maps, that the speed and accuracy of subjects at various spatially relevant tasks are significantly influenced by how they represent space (Hirtle & Jonides, 1985; McNamara et al., 1989; Han & Becker, 2014; Hommel et al., 2000). In addition to helping us understand the influence on cognitive performance, a model of cognitive map structure could facilitate several neighbouring fields, including human-robot interaction (allowing robots to use human-like spatial concepts), artificial intelligence (use insights from human memory to improve artificial memory), and geographic information science (present spatial information in a more easily comprehensible and memorable fashion) - see Section 4.1.

Despite the importance of this question, and perhaps because of the above-mentioned difficulties, no formal theories or models concerning the organizational principles of cognitive maps, able to account for empirical data, have been published since cognitive maps were first proposed to be structured. Little progress has been made on explaining how these representations might be structured in non-trivial, open environments. Multiple features influencing map structure have been suggested, including boundaries in the environment (Wang & Spelke, 2002; Barry et al., 2006), spatial distance and familiarity (Hirtle & Jonides, 1985), action-based and perception-based similarity (Hommel et al., 2000; Hurts, 2008), and functional / semantic similarity (Holding, 1994). However, to the authors' best knowledge, these influences have never been compared based on behavioural data.

A few formal models of map structure do exist - which are predominantly empirically untested -, e.g. the graph-based model by (Thomas & Donikian, 2007) for outdoor virtual reality environments, or based on predicate logic (Reineking et al., 2008), for indoor environments (neither of these have been evaluated against human data); as well as more neuronally plausible but functionally simpler models of place cells such as (Sato & Yamaguchi, 2009) (also empirically untested), or the model by (Byrne et al., 2007) (which can account for lesion effects in humans, but not for large-scale cognitive map structure). Voicu (2003) has published

[1] Some researchers have argued that humans depend on landmark-based instead of map-based navigation whenever they can (Foo et al., 2005), and that most animal behaviour can be explained without the cognitive map hypothesis (Bennett, 1996). However, the well-established body of neuroscientific evidence for dedicated brain regions containing allocentric spatial representations (Moser et al., 2008; Derdikman & Moser, 2010) - both in human and non-human mammals -, together with the ability of human subjects to plan complex novel shortcuts or detours or produce sketch maps, render the idea of allocentric, map-like representations - at least in humans - difficult to dismiss. On the other hand, 'cognitive maps' might well be different from geographical maps in several respects, including being limited in scope, detail, and accuracy, being dynamic, and possibly using metrics that are not (or not exclusively) Euclidean (Spelke et al., 2010; Jeffery, 2015).

the modelling work closest in spirit to the predictive models reported below, utilizing self-organizing maps to model hierarchical cognitive map structure, and reporting that on average, the model exhibits similar distance estimation error patterns to the estimation biases (averaged over all subjects) reported by Hirtle & Jonides (1985). However, this model has not been compared to individual subject maps; and is unable to account for per-subject data in Hirtle's dataset, since it uses only Euclidean spatial distance and no other features (whereas many of Hirtle's subjects do not cluster exclusively based on spatial distance). To date, no empirically tested, formally defined model exists that would be able to predict, or even quantitatively explain, the structure of the individual spatial maps constructed by humans in unconstrained large-scale environments; and the features of the psychological spaces [2] underlying such a model have not yet been explored empirically.

Formulating models and testable hypotheses precisely and unambiguously, is important for efficiently driving research, especially in interdisciplinary areas such as spatial memory (which is of interest in psychology, neuroscience, and artificial intelligence, among other fields). Computational cognitive models are well suited to this challenge as such unambiguous formal descriptions, and provide a common language across disciplines, as well as the additional advantage of very fast prediction generation and hypothesis testing (once the data has been collected, such models can be rapidly run and verified on computers). Thus, they play an important role in the cognitive science of spatial memory, helping to integrate findings, to generate, define, formalize and test hypotheses, and to guide research.

In order to develop and validate a computational model of cognitive map structure, it is necessary - but not sufficient - to tackle the methodological difficulties associated with indirectly inferring consciously inaccessible spatial representation structure from noisy data. In addition, there are also computational challenges. Just like brains can be said to create object representations based on perceived and remembered properties of objects, a computational cognitive model also needs such representations, capturing relevant features. Furthermore, a method is needed that helps to decide which representations should be grouped together on sub-maps. While many low-level features of these representations can be neglected for simplicity in a model on Marr's computational level (Marr & Poggio, 1976), an appropriate metric[3] for capturing similarities between objects is crucially important for exploring how object representations are grouped together onto sub-maps by the brain. Various features, with different levels of importance, can influence whether objects belong together; and defining a metric is a way to formally account for these 'feature importances' (Figure 1). Although the entorhinal cortex has been argued to contain two-dimensional metric grids analogous to graph paper (Jeffery & Burgess, 2006), recent evidence implies the brain's distance metric to be locally distorted (Jeffery, 2015), non-Euclidean (Spelke et al., 2010), and dependent on the above-mentioned features of familiarity, functional similarity, and perceptual similarity (Hommel et al., 2000; Hurts, 2008; Holding, 1994). These features might not be equally important, and their relative importance might not be the same across individuals.

Thus, finding a metric under which objects grouped together by human subjects are 'closer', or more similar, than those not grouped together, is vital for modelling an individual's spatial representation structure. Using this metric, a model can generate predictions regarding which group or sub-map a new object might belong to. Alternatively, objects can be represented in a metric space (which models a subject's psychological space), spanned by relevant features constituting the axes of that space, weighted by their importance. In fact, learning a projection into such a space, in which objects which belong together have a smaller Euclidean distance than those which do not, and learning a metric under which this grouping

[2] By a 'psychological space' we mean a metric space within which the similarities of objects can be represented as distances between points in that space; consistent with the pioneering models of stimulus identification (Shepard, 1957) or categorization (Nosofsky, 1986) or, most recently, conceptual spaces (Gärdenfors, 2004). By a 'feature' of the psychological space, we mean one of the dimensions of this space, which allows measuring similarity along a single aspect, such as the functional similarity of buildings.

[3] A metric (or distance function) is a function that defines a non-negative 'distance' between pairs of objects. Two well-known examples include the Euclidean (geodesic) distance, and the taxicab (Manhattan) distance. In this paper, we model the dissimilarity between two building representations by means of their 'distance' according to a learned metric, which operates on multiple features including but not limited to spatial position (with a distance/dissimilarity value of 0 meaning that the representations are equivalent).

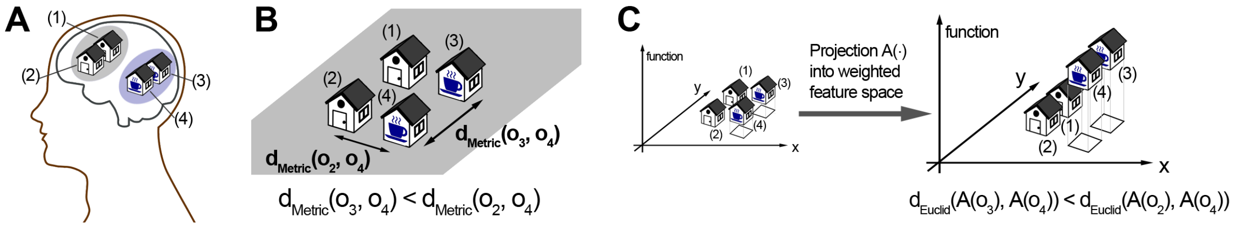

A
(1)
(2)
(3)
(4)
B
(1)
(3)
(2)
(4)
$d_{Metric}(o_2, o_4)$
$d_{Metric}(o_3, o_4)$
$d_{Metric}(o_3, o_4) < d_{Metric}(o_2, o_4)$
C
function
y
x
(1)
(3)
(2)
(4)
Projection A(·) into weighted feature space
function
y
x
(4)
(3)
(1)
(2)
$d_{Euclid}(A(o_3), A(o_4)) < d_{Euclid}(A(o_2), A(o_4))$

Figure 1: Formalizing relative feature importances for grouping objects. Panel A: A subject might group (represent) the two coffee shops together (buildings 3 and 4), even if they are spatially farther apart from each other than to other houses; i.e. (3) and (4) are psychologically closer (more similar) for that individual than (2) and (4). The idea of some features being more important than others when grouping objects can be formally captured either by defining a metric d_{Metric} reflecting the subject's psychological similarity by weighting features appropriately (panel B), or by projecting objects into a feature space (psychological space) spanned by weighted features, in which the Euclidean distances d_{Euclid} are consistent with the subjects' psychological similarity (panel C). The central challenge for a predictive model of spatial groupings is learning these feature importances or 'weights' which parametrize d_{Metric} (or the projection function A) from subjects. Representing subjects' psychological similarities by applying these weights in a distance metric is equivalent in outcome to representing them by functions projecting into a weighted 'psychological space' (similar objects will appear closer under the learned metric / in the learned space).

relationships hold, are two views on the same problem, and the solutions are mathematically analogous (see Figure 1 for an informal and footnote 21 in Section 3.5.1 for a formal argument). In both cases, the solution involves learning parameters corresponding to the relative importances of the features for a given subject. Although traditionally, cognitive psychology has mostly used the former approach, using multidimensional scaling (MDS) to project into a psychological space correctly reflecting similarities (Shepard, 1957), this method is not applicable in our case (the reasons for this are outlined in Section 2.4).

This paper aims to tackle the above-mentioned challenges associated with exploring the structure of spatial representations, and to take a first step towards establishing a formal and empirically substantiated model of this structure. Our core hypothesis is that **the structure of spatial representations in humans arises from a process of clustering** of the represented objects, in a psychological space characterised by multiple relevant information types (features) including the ones mentioned above. Clustering extracts groups or clusters by assuming that objects belonging to the same group (in our case, sub-map) are closer to each other within psychological space than objects belonging to different groups (sub-maps). The characteristics of the psychological space within which this clustering takes place, i.e. which features are relevant and how important they are, has to be learned from participants' responses. The main contributions of this paper are as follows.

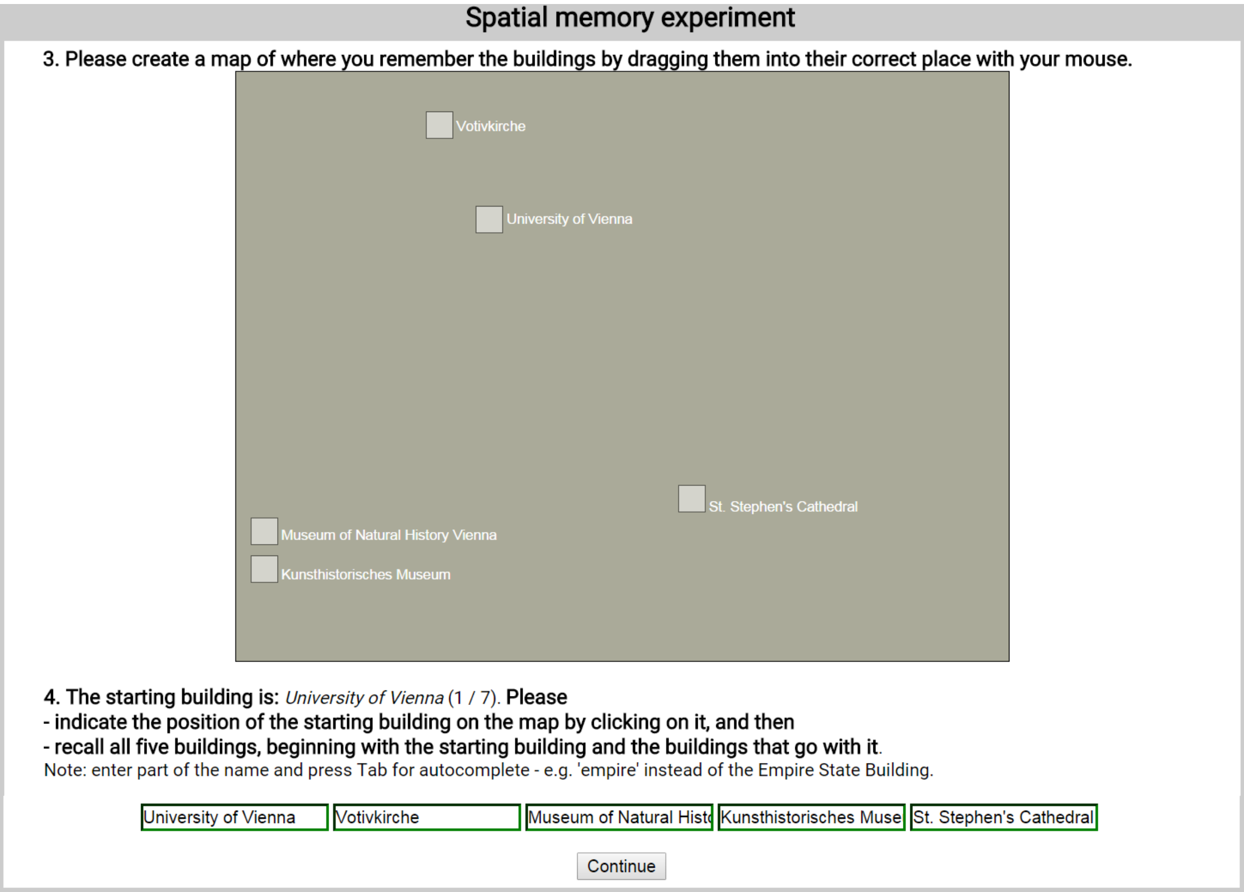
1. We present evidence for the clustering hypothesis both in virtual reality and in real-world environments, and compare the influence of several information types (features) on cognitive map structure, and the stabilities of these feature influences across environments and subjects.
2. We show that the structure of spatial representations, far from being a confounding effect of the recall process or a minor mechanistic detail of memory, has an important role in, and influence on, multiple cognitive phenomena, including (but not limited to) planning, distance estimation, memory accuracy, and response times.
3. We propose and evaluate three computational methods to learn models of subject-specific psychological spaces (either in the form of weighted feature spaces or as distance metrics), even if only small amounts of training data are available
4. We present the first (to our best knowledge) quantitative model able to predict individual cognitive map structures in navigation space, and evidence supporting it.

We only use a few simple types of features for modelling and prediction. Nevertheless, and despite the large amounts of unreliability and noise both in these features and in the participant responses, we show that **spatial map structures can be predicted** for human subjects, in a large number of real and virtual environments. We adopt the behavioural methodology used by (Hirtle & Jonides, 1985; McNamara et al., 1989; McNamara, 1986; Holding, 1994) among others, which infers subjects' representation structure based on recall sequences, and assumes that objects recalled together belong to the same sub-map. Despite of some shortcomings (see Sections 2.1 and 4), the clear and significant influence of the resulting map structures on several kinds of cognitive phenomena (as well as its prior success at showing the influence of hierarchical cognitive maps) lend credence to this method.

Finally, we also make freely available as a web application the experiment software developed to investigate cognitive map structure, at `https://github.com/tmadl/Cognitive-Map-Structure-Experiment`, with the aim to encourage future work on this important but neglected research area.

2. Experimental paradigm

We investigated the structure of spatial representations in navigation space in three experiments. All of the experiments were concerned with the representations of buildings and their relation to each other. In Experiments 1 and 3, subjects recalled real-world buildings that they were already highly familiar with (see Figure 2). In Experiment 2, subjects were presented with three-dimensional virtual reality environments - containing buildings with automatically generated properties - which they had to memorize prior to the recall task from which the representation structure was inferred (see Figure 3).



Spatial memory experiment

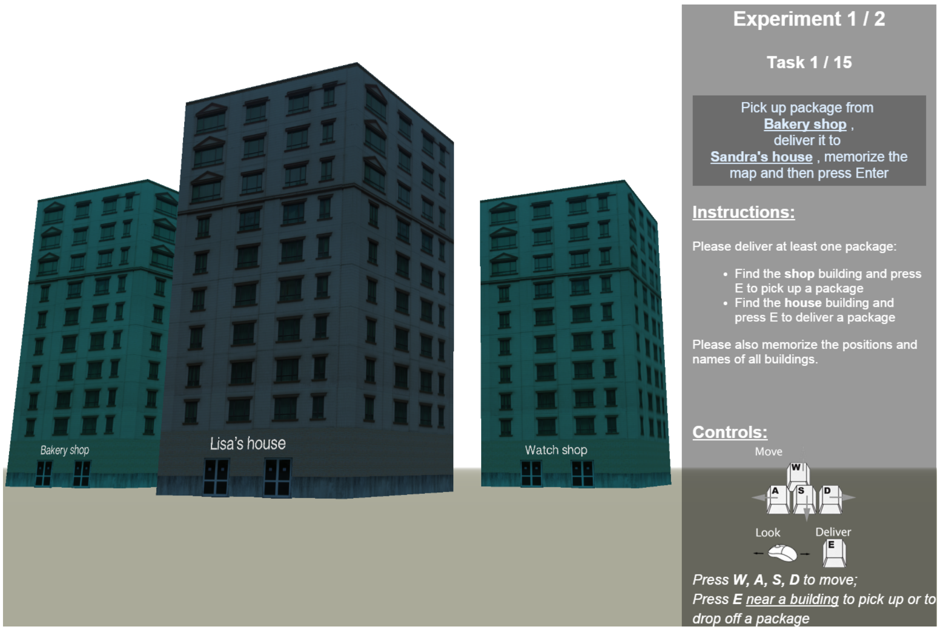
3. Please create a map of where you remember the buildings by dragging them into their correct place with your mouse.

4. The starting building is: *University of Vienna* (1 / 7). Please
- indicate the position of the starting building on the map by clicking on it, and then
- recall all five buildings, beginning with the starting building and the buildings that go with it.

Note: enter part of the name and press Tab for autocomplete - e.g. 'empire' instead of the Empire State Building.

University of Vienna | Votivkirche | Museum of Natural Hist | Kunsthistorisches Muse | St. Stephen's Cathedral

Figure 2: A part of the real-world memories experiment interface of Experiments 1 and 3, with the sketch map question for verifying that subjects have indeed formed allocentric cognitive maps (top), and the recall sequence question requiring them to recall every single building name multiple times (bottom). During this recall question the labelled sketch map was not visible to subjects.



Experiment 1 / 2

Task 1 / 15

Pick up package from **Bakery shop** , deliver it to **Sandra's house** , memorize the map and then press Enter

Instructions:

Please deliver at least one package:

- Find the **shop** building and press E to pick up a package
- Find the **house** building and press E to deliver a package

Please also memorize the positions and names of all buildings.

Controls:

*Press **W, A, S, D** to move;*
*Press **E** near a building to pick up or to drop off a package*

Figure 3: A part of the virtual reality experiment interface of Experiment 2 (the recall sequence interface was equivalent to the real-world experiments; see Figure 2)

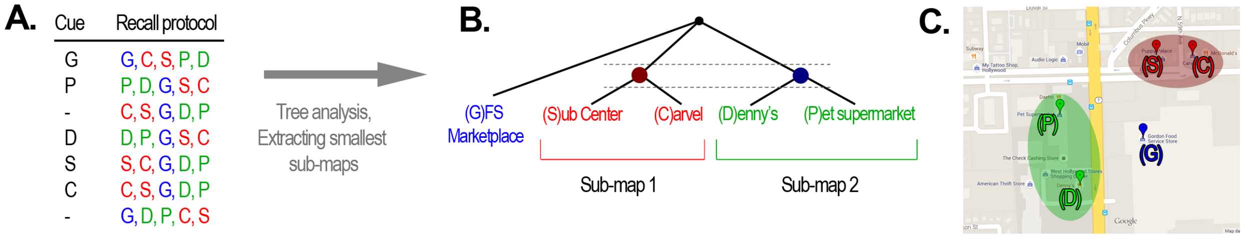
2.1. Extraction of spatial representation structure

To extract the structure of spatial representations, we use a variant of ordered tree analysis on subjects' recall sequences, a behavioural methodology used by (Hirtle & Jonides, 1985; McNamara et al., 1989; McNamara, 1986; Holding, 1994) among others for extracting hierarchies in spatial representations, and by (Naveh-Benjamin et al., 1986; Reitman & Rueter, 1980) for verbal stimuli. The core assumption behind this methodology is that objects recalled together belong to the same representation; i.e. that on the whole, subjects recall every object within a representation (or sub-map) before moving on to the next representation (see Figure 4). Tree analysis operates on a set of recall sequences (with each sequence consisting of all object names, recalled with a particular ordering - usually different from the other recall sequences -, as exemplified in Figure 2A). Variety among these recall sequences is encouraged by cueing subjects with the object they are required to start with (and only uncued parts of the sequence are analysed to avoid the interference of the cue) (Hirtle & Jonides, 1985).

To briefly summarize the collection of these recall sequences (for details, see Section 3.2): in each trial, subjects were first asked to pick a few buildings (5 or 8) within walking distance of each other, which they were very familiar with, and where they knew how to walk from any one building to any other. Subsequently, they were asked to recall the complete list (i.e. recall sequence) of their chosen buildings, starting with a cue building (except for two interspersed uncued trials), multiple times. If building names were missing or incorrect, subjects were prompted again, until they got all of them right. Thus, the ordering within the individual sequences was their only variable aspect.

After obtaining the recall sequences, for each subject, the algorithm simply iterates through all possible combinations of subsets of object names in each recall sequence, finds those subsets which consistently appear together in all sequences (regardless of order), and constructs a hierarchy based on containment relationships from the subsets of items occurring together. The original algorithm also extracts directionality information for each group (whether the items within that group have always been recalled using a consistent ordering). We do not use the order information in the recall sequences in this work (see Supplementary Information for the algorithm we have used). Figure 4 A shows example abbreviated recall sequences, and the resulting tree structure, where each branch or sub-map consists of items which always occur together in the sequences. Unambiguous sub-map memberships are obtained at the level just above the leaf nodes, defining sub-maps as elementary sets of co-occurring items, i.e. those which do not themselves contain further co-occurring items. This procedure partitions buildings into one or two sub-maps in Experiments 1, 2 and 3A, and up to four sub-maps in Experiment 3B.

Since this tree analysis algorithm requires buildings to be recalled together in every single recall sequence in order to infer subjects' sub-maps, it is very sensitive to individual inconsistencies that may result from lapses of attention, task interruptions, and other kinds of noise within participant response (see Section 4 for a discussion and comparison with other approaches of inferring cognitive map structure). To mitigate this, we have eliminated 'outlier' recall sequences, defined as sequences which would have statistically significantly altered the structure if they were included (whereas all others would not). As proposed in previous work on hierarchical cognitive maps (Hirtle & Jonides, 1985; McNamara, 1986; McNamara et al., 1989), we used jackknifing to eliminate outliers. For each sequence, this procedure calculates how the inferred tree structure would change if the sequence were omitted. Trees were quantified using two statistics, tree height and log-cardinality (the logarithm of the number of possible recall sequences consistent with that tree). These statistics were calculated for the tree resulting from all sequences of one trial, as well as for all trees that would result from possible sequence omissions (i.e. if only sequences excluding the omitted one had been entered by the participant). If any of the sequence omissions lead to a statistically significant change in the tree statistics, at a significance level of $\alpha = 0.05$[4], then that sequence was deemed an outlier and was omitted, and the tree resulting from the other sequences of that trial was used for further analysis. All sequences except for outliers were consistent with the same tree structure. Thus, outlier sequences, which significantly changed the tree structure, were likely to arise from the above-mentioned sources of noise (lapses of attention, interruptions, etc.). The outlier sequences detected and removed by the jackknifing procedure comprised 8.5% in Experiment 1, 10.0% in Experiment 2 and 9.5% in Experiment 3, corresponding to less than one omission per subject (across the 7 recall sequences produced per subject and trial).



Cue	Recall protocol
G	G, C, S, P, D
P	P, D, G, S, C
-	C, S, G, D, P
D	D, P, G, S, C
S	S, C, G, D, P
C	C, S, G, D, P
-	G, D, P, C, S

Figure 4: The recall sequence-based method used to extract cognitive map structure. A: Example recall sequences of one of the participants of Experiment 3. Each building was cued once, with two uncued recall trials interspersed (full building names abbreviated by their first character). B: Hierarchical tree structures were constructed by tree analysis, based on the assumption that buildings always recalled together belong to the same sub-map. C: Geographic map of the buildings recalled by this participant. Sub-maps shown in colour, according to the extracted structure.

To simplify the analysis, we subsequently extract the elementary sub-maps (those not containing smaller sub-maps) from the constructed tree - this allows us to model sub-maps, as opposed to full hierarchies. These elementary sub-maps must contain at least two buildings. If a sub-map only contains a single object, then this object is excluded from subsequent analysis. The main reason being that our hypothesis implies sub-maps to be clusters or groups of objects; however, there is no way to verify the plausibility of a single-object cluster (as opposed to clusters containing multiple buildings, for which performance consequences such as between/within-cluster distance biases, priming effects etc. can be investigated - see Section 3.2 for evidence). A further reason for the exclusion of single-object sub-maps is that they were likely to actually be parts of bigger sub-maps in subjects' spatial memories, together with additional buildings not captured

[4] We used a less conservative significance criterion than prior work due to the simpler structures and smaller numbers of objects used (using the extremely conservative significance level of $\alpha = 0.001$ used in the prior work cited above would have led to zero outliers being detected - presumably incorrectly, since it is unlikely that not a single participant would have had any interruptions or lapses of attention).

due to the necessarily limited number of recalled items per trial in our experiments. The exclusion of these single buildings did not have an impact on the plausibility of our claims, since two sub-maps containing pairs of buildings suffice for comparing within sub-map and across sub-map estimations in order to investigate whether map structure has an effect on spatial cognition (see Section 3.2. Experiment 3B collected map structures with eight buildings and up to four sub-maps to show that the model is not limited to two).

A final difference between our methodology and prior uses of the recall order paradigm is the repetition with several different geographic environments for each subject in Experiment 3. Repeatedly extracting cognitive map structures from the same participants is not only interesting, e.g. to compare the variability of the features in subjects' psychological spaces, but also of vital importance for producing and validating a predictive model of the structure of spatial representations. Given the large inter-subject variability in terms of features and feature importances influencing map structure, parametrizing such a predictive model necessitates gathering multiple different cognitive map structures from separate environments (and not just one structure), both for training the model, and for subsequently testing it. The main differences between a repeated and a single-trial paradigm include possible effects of fatigue due to the increased length of experiments, as well as declining accuracy of representations towards the later stages (participants started struggling to cue readily available buildings which they could accurately draw on a map beyond 20 buildings, as evidenced by much slower progress, higher error rates, and much higher rate of participants abandoning the experiment as compared to Experiment 1 which used single trials).

An attempt to mitigate these effects - as well as practical limitations - motivated the decision to use a smaller number of buildings (five in Experiments 1 and 2, five and eight in 3 A and B) compared to the single-trial setup of (Hirtle & Jonides, 1985; McNamara et al., 1989), who used 32 and 28 objects, respectively. Using their dozens buildings for each of the five or three map structures of Experiment 3 would have required participants to recall (and accurately localize) around one hundred buildings or more - as well as judging all of their pairwise similarities, the number of which increases quadratically with the number of buildings (in the case of 32 buildings, they would amount to 496 similarity judgements each for visual and functional similarities, and for each trial, which is nowhere near feasible).

2.2. Experimental platforms and participants

Participants in two of the three Experiments (1 and 3) were recruited from the online survey website Amazon Mechanical Turk (MTurk)[5]. Multiple psychological findings have been replicated before, using subjects from MTurk (Crump et al., 2013), showing the breadth of this platform for psychological experimentation. MTurk offers a participant pool that is significantly more diverse than samples of university students, containing subjects from many countries worldwide and of different age groups; as well being several orders of magnitude larger than most universities' subject pools. But the most important advantage offered by this platform lay in facilitating the collection of information about spatial representations of many, very different geographic environments. Such variety is critical for two main reasons:

- To facilitate generalizable observations (for example, insights from inflexibly planned city areas such as the grid layout of Manhattan might not have been generalizable to other street layouts), and
- To avoid local biases (for example, using exclusively local maps in the same city for each participant might have led to conclusions about the spatial structure of the local city, reflected in subjects' representations, as opposed to insights into the way subjects structure space in general).

Our objective of collecting cognitive map structures from a large variety of different geographical environments was indeed successful - we collected data and analysed spatial representations from several environments within **149 different cities** across multiple continents (see Figure 5 - a list of these cities can be found in the Supplementary Information).

[5] https://www.mturk.com

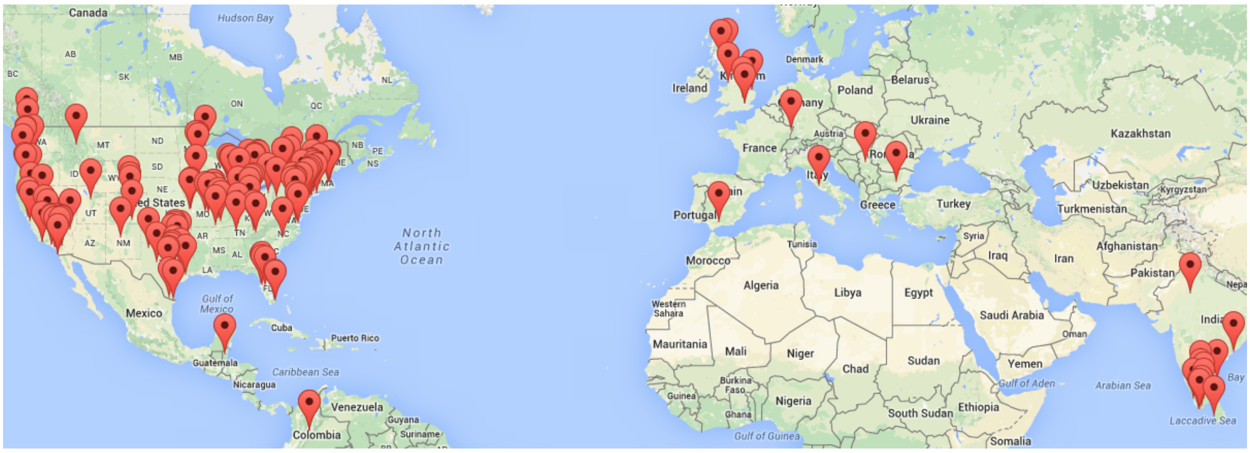


Figure 5: Overview over the 149 cities in which participants' spatial memory structures were extracted (and predicted by the computational model) in the real-world experiments (full list in Supplementary Information)

2.3. Exclusion of participant maps not significantly better than random chance

Throughout this paper, we have only analysed participants' data if their sketch maps were significantly better than random chance, in order to avoid false conclusions about cognitive maps being made on the basis of non-allocentric representations. Since route knowledge suffices for navigating between buildings, participants might have lacked survey knowledge about some of the buildings in these experiments. To rule out participant data not showing evidence of allocentric cognitive maps, we first performed a test of participants' sketch maps against randomness, before carrying out the subsequent analyses described in Section 3.

We compared the sum of squared errors (SSE) calculated by subtracting the positions of buildings on participants' sketch maps from those on the correct geographical map (obtained from Google Maps), with the SSEs of 10,000 randomly generated maps containing 5 buildings against the correct map. Since subjects' sketch maps were produced on empty surfaces without any position, orientation or scale cues, as seen in Figure 2, they were first aligned (translated, rotated, and scaled) with the correct map using Procrustes analysis (Gower, 1975) without reflection. The randomly generated maps were also aligned in the same fashion. The distribution of the SSEs of 10,000 [6] Procrustes-aligned random maps was then used to test whether subject maps were better than random. Specifically, subject SSEs were tested against the null hypothesis that they were drawn from the distribution of uniformly random map SSEs. Two different significance tests were applied at $\alpha = 0.05$ significance level, and found to largely agree (in all but 3% of the cases in Experiment 1, 1.3% in Exp. 2 and 4% in Exp. 3): a Z-test assuming normal distributions of SSEs, and a non-parametric Bootstrap hypothesis test (MacKinnon, 2009), which requires subject maps to be better than a proportion of $1 - \alpha$ of the random maps.

Because the former test makes the assumption of normally distributed data, which is incorrect for the vast majority of distributions of random map SSEs according to Shapiro-Wilk tests, we use the latter, non-parametric hypothesis testing method throughout this paper to test participant maps against randomness.

2.4. Data analysis

Subject data collected using the recall order paradigm was analysed as follows. Maps not significantly better than random (and corresponding recall lists), and recall sequences not containing structure (where no items consistently occur together), were not analysed further, since the former show no evidence of allocentric representations of the buildings on that map being present in the subjects' spatial memory, and the latter

[6] We have also tried higher numbers or randomly generated maps, and found that 10,000 samples suffice for an approximation of the distribution under the null hypothesis, since increasing this number to 15,000 or 20,000 did not make a difference.

shows no structure to be analysed. Next, map structures (sub-map memberships) were derived using tree analysis, and pairwise distances in all features were calculated (see Section 3.3). This data allows reporting the influence of features on map structure, and the inter- and intra-subject variability of this influence (Section 2.2). It also allows inferring subject-specific models reflecting the individual feature importances of a subject, in the form of a metric or a 'psychological space' (subject-specific feature space) - see Figure 1. A clustering algorithm (described in Section 3.4), operating on the learned model, allows prediction of sub-map structure, if the clustering hypothesis is plausible (assuming that cluster memberships correspond to sub-map memberships).

Simple clustering based on a Euclidean distance metric, in the original feature space, fails to account for participant map structures. The main reason for this is that different participants might not rely on the same set of features; and the relative importance of these features might also differ across subjects (see below, particularly Figure 7, for evidence). For this reason, learning an appropriate subject-specific model (metric, or feature space) is crucial in order for our computational model to provide accurate predictions. Since learning human representation similarity metrics (or psychological spaces) from highly noisy and sparse data is a largely unexplored problem in the cognitive sciences[7], we turn to machine learning for possible solutions. We describe and empirically test three computational methods for tackling this problem below (see Sections 3.4.1 and 3.5.1, as well as the Supplementary Information for details).

Together, these learned subject-specific models, and the clustering algorithm, constitute a computational model of cognitive map structure learning able to predict sub-map structure in advance, and allowing the verification of such predictions (Sections 3.4 and 3.5 compare the model predictions against human data).

3. Experiments

3.1. Overview of the experiments

This section reports the results of four experiments investigating the principles underlying cognitive map structure. Experiment 1 (Section 3.2) is concerned with the question of whether this kind of structure uncovered by the recall order paradigm is relevant - whether it impacts cognitive performance in other ways than recall sequences -, investigating effects on distance estimation biases, sketch map accuracies, estimated walking times, and planning times in real-world environments well known to subjects.

The plausibility of our central hypothesis - that cognitive map structure arises from clustering - is investigated in the subsequent section (3.3), also in real-world environments chosen by subjects themselves. This claim requires buildings that are more similar (closer in psychological space) to be more likely to be grouped together in long-term memory representations, and thus more likely to be recalled together. We report correlations between the probability of buildings being represented together, and proximity in various features relevant to cognitive mapping. We also make between- and across-subject comparisons with regard to feature importances.

Since a good model should be able to make predictions, we proceed to report the predictability of spatial representation structure. We use a clustering model in order to predict map structure, assuming that cluster membership in an appropriate 'psychological space' (i.e. weighted feature space) corresponds to sub-map memberships.

However, a model clustering buildings in a static feature space fails to produce accurate predictions, simply because there exists no feature space generalizable across participants (see Section 3.3). In order to learn subject-specific feature spaces, we utilize three methods to uncover the features and feature importances spanning the psychological space hypothesized to underlie spatial representation grouping in Experiments 2 and 3. We collect map structures of several different environments from the same subjects in these experiments, using a subset of them to learn a model, and testing its predictions on the remaining subset.

[7] There has been an approach used in cognitive psychology for projecting data into a space in which distances reflect subject similarities, called multidimensional scaling (MDS) (Shepard, 1957). This method is not applicable in our case, because it requires a full pairwise distance matrix. However, our training data comes from several different environments; and pairwise distances and similarities are only known within, and not across, those environments.

In Experiment 2 (Section 3.4), subjects are asked to learn spatial memories of 3D virtual reality environments. Unlike the other experiments, this approach allows full control over all properties of the stimuli being memorized. Utilizing this flexibility of virtual reality, we report prediction results using clustering, and the decision hyperplane method for learning subject-specific models, which tackles the challenge of inferring multiple feature importances from few data points by generating the environments such that participants' subsequent responses minimize the uncertainty of the model regarding the feature importances, inspired by active learning in machine learning (Settles, 2010). After subjects have been queried on a reasonable number of environments, and the model's uncertainty regarding their psychological space has decreased, they are presented with completely random environments, on which the trained models are tested. We report prediction accuracies both on environments generated such that they minimize model uncertainty (using active learning), and on random environments.

Although virtual reality allows fine-grained control over memorized environments, it is necessarily composed of strongly simplified stimuli and less complex surroundings. To show that the approach of inferring subject-specific models and subsequently clustering objects can also successfully predict cognitive map structure in the much more complex real world, we once again collect data from subjects' spatial memories of real environments freely chosen by them in Experiment 3 (Section 3.5). Since the approach of optimally minimizing model uncertainty is infeasible when using uncontrolled real-world memories, we use two more general methods to infer subjects' psychological spaces, global optimization and Gaussian Discriminant Analysis (see Section 3.5.1). Of these, the latter is novel, and to our knowledge the only metric learning approach applicable to our data. We report prediction results on data excluded from the model training process, substantiating our central hypothesis, and showing, for the first time, the predictability of spatial representation structures on the individual level.

3.2. Experiment 1 - Relevance of cognitive map structure extracted from recall sequences

This experiment was conducted to substantiate the recall order paradigm used throughout this paper to infer cognitive map structure. If this paradigm infers something about actual representation structures in spatial memory, then the uncovered structures should have a significant impact on both the speed and accuracy of memory recall for spatially relevant information. To avoid possibly confounding effects of stimulus presentation and memorization, the stimuli used were ones participants had already committed to their long-term spatial memory - the experiment used buildings subjects were already very familiar with and could easily recall information about[8].

Although data consistent with two of the results presented in this section (the effects of map structure on distance estimation biases and sketch map accuracies) have been observed and published before, this prior work had used significantly fewer subjects than our experiment, and exclusively university students, unlike our participants. (Hirtle & Jonides, 1985) had six participants, reporting distance biases and sketch map accuracies; and (McNamara et al., 1989) had twenty eight, reporting only the former.

3.2.1. Participants

One hundred and fifty participants were recruited, consented, and compensated through the Amazon Mechanical Turk (MTurk) online survey system (78 females, 74 males). Participants were required to have at least 95% approval rating on previous MTurk jobs to ensure higher data quality, and all of them were over 18 years of age (as required by the website).

3.2.2. Procedure

The experiment was conducted on a website participants could access through MTurk after giving their consent. In the first two questions, subjects were asked to enter the name of a city they were very familiar

[8] The possible objection that the structures might be induced by the experimental paradigm, and learned by participants during the trials, can be excluded, because of the approximately uniform distribution of the outlier sequences (the first few sequences were not more likely to be outliers than the last few sequences, and no evidence for any learning of map structures during the real-world experiments could be found in the data - see Supplementary Information for details).

with, and, subsequently, to pick five buildings they know well. Thus, well-established long-term memories were tested instead of novel stimuli. Subjects were instructed to make sure that they knew where in the city these buildings were located, how to walk from any one building to any of the others, what each building looked like, and what purpose it served. They were only able to proceed past this stage if the website was able to locate all five of the buildings on a geographical map (Google Maps API[9] was used to look up the latitude and longitude of each building).

To verify that subjects had indeed formed allocentric spatial representations of the area of the city they had selected, and to allow the analysis of the accuracy of their representations, they were also asked to produce a 'sketch map', by dragging and dropping five labelled squares representing their buildings into their correct place using their mouse (Figure 4A, top). No cues or information was provided on the sketch map canvas, just an empty gray surface with five squares labelled according to subjects' entered building names. Thus, only the relative configuration of the buildings was analysed in this research, after optimal translation, rotation and scaling to fit the placement and size of the correct map as well as possible, by using Procrustes transformation (Gower, 1975).

After the sketch map, subjects performed a seven-trial recall test. In five of the seven trials, they were given a cue or starting building, and were instructed to *'recall all five buildings, beginning with the starting buildings and the buildings that you think go with it'*, encouraging recall of building names in the order they came to subjects' mind, closely following the instructions given by (Hirtle & Jonides, 1985; McNamara et al., 1989) and others. In the remaining two, uncued trials, subjects were asked to start with any building they wished. If subjects omitted or incorrectly recalled any of the buildings, they had to repeat the trial (thus, only the ordering changed across trials).

The recall test allowed the experiment software to immediately infer subjects' map structure using the tree analysis algorithm (see Section 2.1. Smallest sub-maps - those not containing further sub-maps - were extracted). The next stage of the experiment proceeded based on this structure. Participants were first asked to estimate the time required to walk between four pairs of buildings. Unbeknownst to them, two of the estimations concerned within-, and two of them across-sub-map pairs, in randomized order, and were generated such that the Euclidean distances in the within-cluster trials were as close as possible to the distances in the across-cluster trials, to mitigate effects of simple distance, as opposed to map structure. After reading the instructions in their own time, subjects were told to estimate and enter the walking time in minutes (the time required to walk from one of these buildings to another) as rapidly as possible. Their responses, as well as their response times (time elapsed between presentation of the pair of buildings for walking time estimation and subjects entering a number and clicking a button) were recorded.

In a subsequent stage, also based on the uncovered map structure, participants had to estimate the distance between four pairs of buildings (Euclidean distance - 'as the crow flies' - as opposed to the walking times of the previous stage). Once again, two within-cluster and two across-cluster pairs were selected such that within- and across-cluster trials differed as little as possible from each other in terms of spatial distance.

Finally, once again in an untimed fashion, subjects were asked to judge the similarities of all pairs of buildings, i.e. $\binom{5}{2} = 10$ pairs, as well as a control pair of one of the buildings to itself, both in terms of visual similarity, and similarity of purpose/function - thus, they had to enter 2x11 similarity judgements. Similarities were judged with the help of 1-10 rating scales, with 1 standing for not similar and 10 for very similar. The two self-similarity judgements were randomly interspersed and verified to avoid subjects rushing the process or entering random values.

Ground truth geographical maps containing participants' self-chosen buildings were constructed by obtaining latitude and longitude coordinates from Google Maps API, and utilizing an elliptical Mercator projection to obtain x and y coordinates suitable for comparison with subjects' sketch maps. Euclidean distances between buildings were also calculated based on this projection (as this procedure is more accurate than most alternatives such as the Haversine formula). Finally, path distances as well as ground truth walking times were obtained from Google Directions API [10], which plans the shortest possible walking route

[9] https://developers.google.com/maps/

[10] https://developers.google.com/maps/documentation/directions/

	Actual distance (m)	**Estimated distance (m)**	**Distance bias (Estimated-Actual)**	**Estimated walking time (min:sec)**	**Response time when estimating walking time (s)**
Within mean Within std	$\mu = 1242$, $\sigma = 1508$	$\mu = 676$, $\sigma = 1036$	$\mu = -574$, $\sigma = 1825$	$\mu = 8:43$, $\sigma = 8:23$	$\mu = 8.4$, $\sigma = 6.0$
Across mean Across std	$\mu = 1245$, $\sigma = 1931$	$\mu = 1139$, $\sigma = 1739$	$\mu = -146$, $\sigma = 1703$	$\mu = 12:45$, $\sigma = 11:36$	$\mu = 18.0$, $\sigma = 92.3$
Significance of difference	$p = 0.109$ (nonsignificant), $U = 12594$	$p = 0.019$ (significant), $U = 12502$	$p = 0.047$ (significant), $U = 11900$	$p = 0.001$ (significant), $U = 11009$	$p = 0.030$ (significant), $U = 13097$

Table 1: Effects of spatial representation structure on distance estimation, walking time estimation, and response times. All of these estimated magnitudes, as well as response times, are significantly smaller when both buildings are on the same sub-map (i.e. on the same representation) compared to when they are not. Data from 380 pairs of buildings were compared (269 across sub-maps, and 111 within sub-map). Apart from the representation-dependent biases, subject estimations were reasonably accurate (correlation of $r = 0.40$ between estimated and actual Euclidean distance, and $r = 0.48$ between estimated and actual walking time as calculated by Google Maps)

between two buildings along pedestrian paths (which is usually distinct from, and longer than, Euclidean or 'beeline' distance).

3.2.3. Results

Participants with sketch maps not significantly better than random chance were excluded (using the procedure described in Section 2.3). 86 participants with reasonably accurate survey knowledge of their chosen environments remained (40 female, 46 male). Of these participants, 53 had structure apparent in their recall sequences (20 female, 33 male). The difference in the ratio of structured representations between male (72%) and female (50%) participants is statistically significant at $p = 0.04$ ($U = 4.39$) according to a Mann-Whitney U test. We employed this test here and for a majority of our other significance tests (unless otherwise specified), because the tested variables were not normally distributed according to a Shapiro-Wilk normality test ($p = 0.00$, $W = 0.63$), violating the assumptions behind ANOVA or t-testing. The Mann-Whitney test is a nonparametric test which has greater efficiency than the t-test on non-normal distributions (and is comparably efficient to the t-test even on normal distributions) (Nachar, 2008).

To test whether map structure has an impact on other cognitive phenomena, we compared estimations of distance, walking times, and planning times, between pairs of buildings lying on the same representation (within sub-map estimations), and pairs of buildings on different representations (across sub-map estimations). Table 1 reports the results (6 across sub-map and 1 within sub-map distance estimations were excluded, because they exceeded $10km$, clearly violating the instruction of being within walking distance). Reported correlations are Spearman's correlation coefficients, here as well as throughout the paper.

In order to avoid effects arising purely from differences in spatial distance, we have queried subjects on the pairs of their buildings (among all possible pairs) which were the least different in spatial distance. In these comparisons, effects purely of spatial distance are unlikely, since distances were not significantly different between within sub-map and across sub-maps estimations ($1242m$ and $1245m$ on average) - according to a Mann-Whitney test ($U = 12594$, $p = 0.11$), the difference is not significant.

We have also examined the effect of whether maps were structured on sketch map accuracies. The sum of squared errors (SSE) between the resulting sketch map building positions and the geographical building positions were calculated, and SSEs for all maps with structure ($\mu = 0.305$, $\sigma = 0.276$) were compared to the SSEs for maps without structure ($\mu = 0.370$, $\sigma = 0.307$). SSEs were found to be significantly smaller for structured than for unstructured maps ($p = 0.019$, $U = 2325$), hinting at a correlation between map accuracy and structuredness which can indeed be observed ($r = -0.17$, $p = 0.04$).

Finally, the SSEs between sketch map and geographic map distances were compared for pairs of within sub-maps and pairs across sub-maps, after alignment and normalization. The sketch map distance SSEs within sub-maps ($\mu = 0.607$, $\sigma = 1.677$) were significantly smaller than those across sub-maps ($\mu = 0.916$, $\sigma = 1.53$) according to a Mann-Whitney U test ($p = 0.023$, $U = 6304000$).

3.2.4. Discussion

The highly significant differences in the accuracies of sketch maps, distance and walking time estimations, which all depend on whether or not the buildings involved in the estimation are on the same sub-map or on different sub-maps, substantiate the claim that the structures uncovered by this method are indeed relevant, and play a significant role in multiple cognitive mechanisms.

The trends in the distance error biases - distances generally being underestimated within sub-maps compared to across sub-map estimates - match previously made observations using smaller numbers of subjects (Hirtle & Jonides, 1985). The main difference is that this previous work has found underestimation within- and overestimation across sub-maps, whereas our results suggest underestimation in both cases. The negativity (underestimation) of the across sub-map distance estimation errors is statistically significant compared to the null hypothesis that there is zero or positive bias ($p = 0.03$, $U = 4937$).

Both the difference in estimated walking times, and the differences in the response time in this question, are novel results. As opposed to Euclidean distance estimation or sketch map drawing, which can be done by glancing at or recalling a geographical map, accurate walking times are difficult to estimate without actually having explored this environment and being able to plan the routes in question. Subjects need to mentally plan the route and simulate the walk to estimate the time (or to recall the duration of the walk from long-term memory, should the durations of all walks between all possible building pairs be readily memorized by subjects, which is unlikely). The observation that the mean time required to do so more than doubles across sub-maps, compared to within (and that the variance in RTs increases by an order of magnitude) provides additional, substantial evidence for the relevance of map structures - as inferred from recall sequences - to spatial cognitive processes.

3.3. Clustering and features determining map structure

In the Introduction, we have hypothesized that the structure of spatial representations in humans arises from clustering within some psychological space. In this section, we investigate the plausibility of this hypothesis. If this was the case, we would expect the probability of a pair of buildings being co-represented (i.e. represented on the same sub-map) to strongly depend on their ‘similarity’ or distance across various features including spatial distance, with stronger dependencies for spatially relevant features compared to semantic or visual features. We would also expect several such features to play a role, since distance alone is insufficient to explain previous results (Hirtle & Jonides, 1985; McNamara et al., 1989). We would expect the relevance of each feature to be apparent from its influence on map structure, measurable by the correlation between co-representation probability (the probability that two buildings are co-represented on the same sub-map) and the distance in this feature. Finally, we would expect large inter- but small intra-subject variability in these correlations, i.e. stable feature relevances within subjects which are not necessarily generalizable across subjects, analogously to psychological spaces for concept representation (Nosofsky, 1986; Gärdenfors, 2004).

We investigate several features listed below, motivated by hints in the literature that they might play a role in the representation structure of object-location memory.

1. Remembered distance, i.e. the distance on subjects’ sketch maps
2. True Euclidean distance based on geographical maps
3. Path distance (or ‘city-block’ / ‘Manhattan’ distance), since recent brain imaging evidence suggests that the hippocampus - a spatially relevant brain region - represents both Euclidean and path distances (Howard et al., 2014)
4. Boundaries in the environment (such as rivers, cliffs, city walls, etc.) - based on neuroscientific and behavioral evidence that boundaries play an important role in spatial memories (Wang & Spelke, 2002; Barry et al., 2006)

5. The number of streets separating a pair of buildings (intersecting with a straight line connecting these buildings)
6. The sizes of separating streets; that is, whether these streets could easily be crossed (whether or not they were highways/motorways/primary roads which are difficult for pedestrians to cross)
7. Visual similarity (as indicated by participants), since clustering by perceptual properties has been reported (Hommel et al., 2000), and vision has been suggested to be vital to spatial representation (Ekstrom, 2015),
8. Functional similarity, or similarity of purpose, as indicated by participants - because action-based similarity has been claimed to have an effect on spatial memory (Hurts, 2008), and also because of the importance of action-related roles within the influential grounded cognition paradigm (Barsalou, 2008).
9. Phonetic [11] and morphological [12] similarity of building names. The main motivation behind including these features was to investigate any possible interference on the structures inferred from recall sequences caused by verbal short-term representations. Subjects might employ some short-term representation strategy to complete the recall trials more rapidly - instead of recalling from long-term spatial memory -, for example subvocal rehearsal loops (articulatory loops). Including phonetic and morphological similarity features helps measure the effect of such verbal strategies.

The first six of these features - remembered, Euclidean and path distance, and boundaries, separating streets, and crossable streets - were obtained based on geographical data available online. Most such geospatial ground truth data used was obtained using Google's publicly available Maps API, with the exception of boundaries in the environment, and crossable streets (whether separating streets were difficult to cross) - these two features were obtained from Open Street Maps (OSM) through their publicly available API called Overpass [13]. As in all experiments in this paper, ground truth maps and distances are based on an elliptical Mercator projection of latitudes and longitudes obtained from Google Maps API, except for path distances and walking times which were queried from Google Directions API.

All features were converted into distances / dissimilarities before subsequent analysis. Similarity features, such as visual, functional, phonetic and morphological similarities, were subtracted from the maximum value possible for that feature to obtain corresponding dissimilarities.

3.3.1. Participants, Materials, and Procedure

Data from Experiment 1 (Section 3.2 above) as well as Exp. 2 (see Section 3.4) and Exp. 3 A and B (see Section 3.5) were analysed with regard to the plausibility of the clustering hypothesis, as well as the underlying features determining map structure. Thus, the participants, materials and procedures for data collection were exactly the same as in those experiments, following the recall order paradigm described in Section 2.

All Figures in this Section are split into four parts, for Experiment 1, Exp. 2, and conditions A and B of Experiment 3. We report results separately, since there were slight changes in procedure. Briefly, Exp. 2 was conducted in three-dimensional virtual reality environments, whereas the other experiments used subjects' established real-world spatial memories. Furthermore, cues were presented verbally in Exp. 1 and 2 and spatially, highlighted on sketch maps, in Exp. 3 (Exp. 3B also used 8 buildings, unlike the 5 used in the other experiments). Finally, Experiments 2 and 3 tested spatial memories of several different environments, in order to facilitate learning a model and testing predictions, whereas Exp. 1 did not.

[11] Phonetic similarities have been determined using the Double Metaphone (Philips, 2000) phonetic encoding algorithm, since it is more accurate than older alternatives such as Soundex, and also accounts for a large number of irregularities in multiple languages, not just English (vital since many participants from several non-English speaking countries participated in this experiment - see Supplementary Information for a detailed list). For building names consisting of multiple words, the sum of the phonetic similarities of the constituent words was used.

[12] Morphological similarities were calculated based on recent work by (Khorsi, 2013), implemented by the PhonologicalCorpusTools library accessible at http://kchall.github.io/CorpusTools/. For building names consisting of multiple words, the sum of the morphological similarities of the constituent words was used.

[13] http://overpass-api.de/

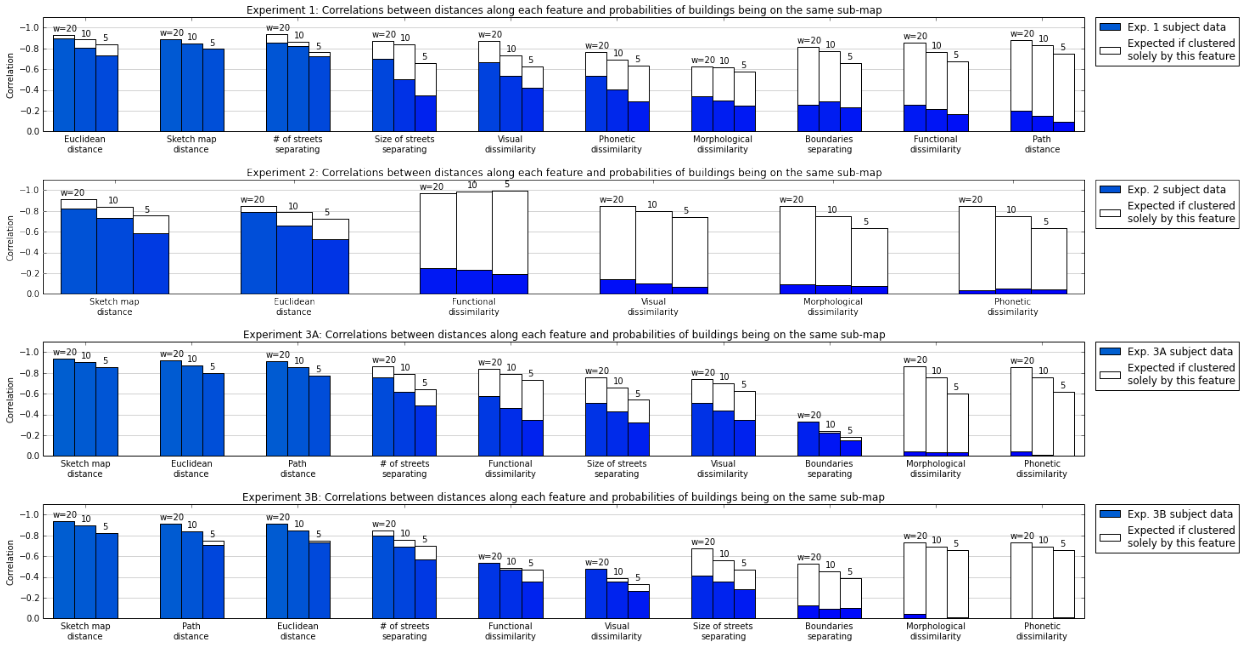


Figure 6: Correlations between probabilities of being on the same sub-map, and distances along each feature, for pairs of buildings in Experiments (from top to bottom): 1, Experiment 2 in virtual reality (therefore lacking geospatial features), and 3A, 3B. Correlations are reported separately for each feature. The three bars per feature show results at three different window sizes w used for calculating co-representation probabilities (higher w lead to less noisy probability estimates through smoothing, resulting in higher correlations). Empty bars show levels of correlation that would be expected if maps were clustered according to the single respective feature only.

3.3.2. Results

The clustering hypothesis introduced in Section 1 implies that buildings closer together in psychological space are more likely to be represented on the same sub-map in participants' spatial memory. To test this hypothesis, we investigated the correlations between the probabilities of pairs of buildings belonging on the same sub-maps and between the distance between them, along the various features listed above.

Figure 6 provides an overview of the correlations of these features with the probabilities of co-representation on the same sub-map. These probabilities were calculated using a moving average with window w of the binary vector indicating whether or not pairs were stored on the same sub-map - simply put, the likelihood of co-representation at a specific distance equals the ratio of the number of co-represented pairs divided by the number of all pairs within some small window w close to this distance (for example, if $w = 3$ and out of three building pairs with distances $95m$, $100m$ and $105m$ two were represented on the same sub-map, then the probability of co-representation at $100m$ would equal $p = 2/3$).

The Figure also shows the correlations that could be expected if participant's map structure had arisen from clustering by just that one feature (empty bars in Figure 6) - i.e. the correlations that would have been observed had participants 1) used clustering to structure their maps, and 2) used only distances within one respective feature for this clustering. These expected correlations were calculated using the same participant data; but artificially structuring the subject map - using clustering along one respective feature - instead of using subjects' sub-map memberships. Gaussian mixture models (GMMs) (Redner & Walker, 1984) were used for the artificial structuring, just like for prediction in the computational models described below, since they are more psychologically motivated than other clustering algorithms (see Section 3.4.1).

Next, we have investigated the variability of the reliance of these features within and across subjects; i.e. whether the same features were used by - and whether they were similarly important for - all subjects, and whether they were the same for individual participants in different environments. Figure 7 shows the standard deviations of the co-representation correlations of these features, across subjects (left panels)

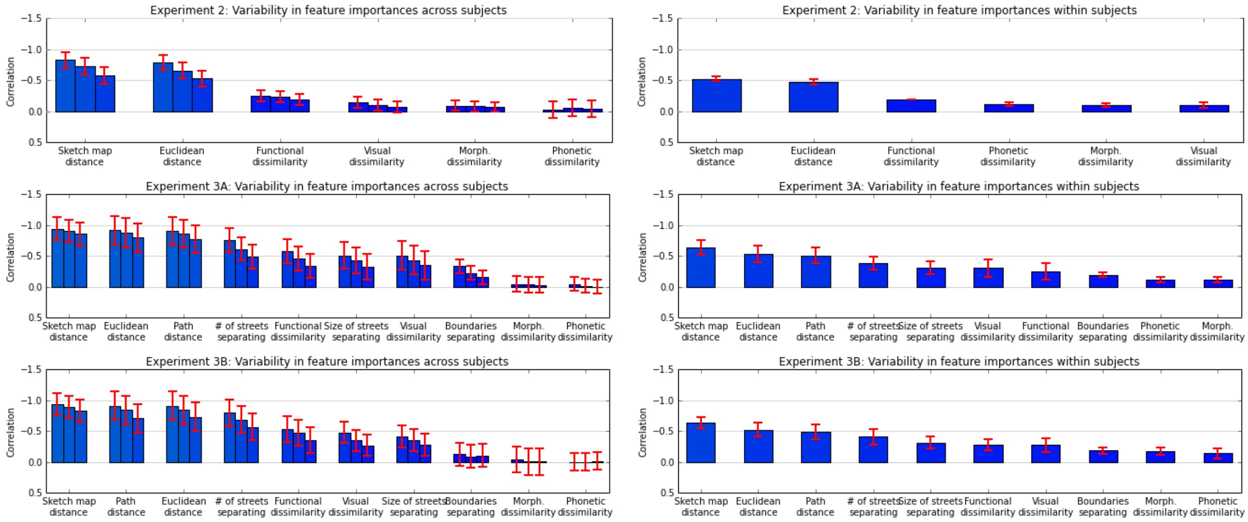


Figure 7: Variability of features influencing cognitive map structure. Feature variabilities across all subjects (left) and across map structures of individual subjects (right) are shown, plotted as error bars on each average feature correlation. Top: Bottom: Feature variabilities in the test trials of Experiment 2. Middle: Feature variabilities in Experiment 3A. Bottom: Feature variabilities in Experiment 3B.

and within subjects, i.e. across the maps of individual subjects (right panels), averaged over all subjects. Specifically, the standard deviations of the point biserial correlation coefficients [14] between the feature distances and co-representation probabilities are reported for all error bars in the plot. For the within-subject plots (right panels), the magnitude of the bars is also calculated using point biserial correlation, for the same reason - there being too few within-subject building pairs for the moving average-based probability calculation.

Finally, according to Shapiro-Wilk normality tests, none of the distributions of feature correlations are normally distributed (all p values for all features are many orders of magnitude less than 0.01). Rather than most subject structures arising from these features weighted in the same fashion, or from feature correlations concentrated around a most common value, the particular strengths of the influences on participants' representation structures seem to be irregularly distributed. However, they are significantly more consistent across individual participants' map structures (Figure 7 right) than across all participants (Figure 7 left).

3.3.3. Discussion

The strong dependence of co-representation probability on distance along various features (Figure 6) provides strong evidence for the plausibility of the clustering hypothesis. Furthermore, confirming intuitive expectation, spatial features show a much stronger influence on map structure than other, non-spatial dimensions.

Figure 7 shows that there is a large amount of variability in the importance of different features to various subjects. This spread is significantly less across the map structures of individual participants (Figure 7 right) than across all participants. Thus, although collecting a high enough number of map structures to reliably infer subject-specific feature importances presents several practical challenges (see next section), doing so is unavoidable for predicting spatial representation structures.

It is important to point out that correlation with co-representation probability alone is not a sufficient metric for describing the influence exerted by a feature on cognitive maps. There might be indirect causation

[14] Biserial correlation with the binary vector indicating same or different sub-map pairs was used, instead of calculating probabilities and using continuous correlation, because the numbers of available within and across sub-map pairs of buildings for a specific map of a specific participant were frequently below the window sizes used for estimating co-representation probabilities in Figure 6.

or a common cause, or deceptively low correlations due to sparse data (for example, very few natural boundaries are present in most cities, which causes low correlations despite their importance according to the results below), or other reasons for correlation not translating to causation.

For this reason, to what extent different features facilitate the prediction of individual map structures is a more meaningful measure of their importance in the cognitive map structuring process. The following sections report prediction results, both in automatically generated virtual reality environments (Section 2) and in real-world environments freely chosen by subjects (Experiment 3).

3.4. Experiment 2 - Predictability of map structure in virtual reality environments

This experiment investigated the question whether the clustering hypothesis allows robust advance prediction of participant map structures. Because of the observation that feature importances vary greatly across subjects, but less for individuals (Figure 7), it was designed to first learn these per-subject importances, before producing predictions using a clustering mechanism. This process was inspired by *active learning* (Settles, 2010), a field in machine learning which allows algorithms to choose the data from which they learn, thus facilitating better performance with less training data. This latter point is crucial for our experimental paradigm - as inferring the representation structure of even small environments with few buildings requires several full recall sequences, there is a practical limit on how many structures per participant can be produced - thus, this limited budget of data should be used in a fashion close to the statistical optimum. Optimally reducing model uncertainty using active learning is one possible approach towards this objective.

3.4.1. Computational methods

As described in the Introduction, a computational model of cognitive map structure requires learning subject-specific models reflecting feature importances, as well as a clustering algorithm. For this experiment, which allows full control over the memorized environments, we have used the decision hyperplane method to infer learn feature importances. We constructed a training environment for each trial such that 1) they contained two clusters (shop buildings and house buildings), 2) only the features of a single building, which lay somewhere between the two clusters, were varied (see Figure 8). We trained a linear classifier to assign the middle buildings of all trials of a participant to one or the other cluster in feature space. The class label (dependent variable) y was derived from that participant's recall sequences in each trial ($y = 1$ if the middle building was co-represented - i.e. recalled together - with the shop buildings, and 0 if it was co-represented with the house buildings). The distances of the middle building from the shop buildings along all features (in unweighted feature space) served as predictor (independent) variables x.

Based on these variables, a linear 'decision hyperplane' was calculated, which separated the set of all data points characterizing the middle buildings of a participant's trials into two sets: into middle buildings which were represented together with shops (if below the decision hyperplane) and into those which were co-represented with houses (if above the decision hyperplane) - see Figure 8. The slope of this 'decision hyperplane' in each feature dimension (distance, visual similarity / colour similarity, functional similarity) thus indicated the importance of each feature to this participant (for example, if the decision hyperplane in Figure 8 was horizontal, that would mean that the y-axis - spatial distance - would be the only feature of relevance for this subject. Conversely, if the plane was almost vertical, spatial distance would be unimportant).

The decision hyperplane was calculated using logistic regression (Hosmer & Lemeshow, 2004), formulating the question whether to group the middle building on the shop sub-map or the house sub-map as a binary classification problem. Thus, the probability $P(S|D)$ of clustering the middle building to the shop sub-map, given a set of distances $D = (d_s, d_f, d_p, ..., d_n)$ from the shop buildings along a number of features, including spatial (d_s), functional (d_f) and perceptual (d_p) distance (difference in colour), was modelled using the logistic regression equation

$$P(S|D) = \frac{1}{1 + e^{-W^T D}}, \tag{1}$$

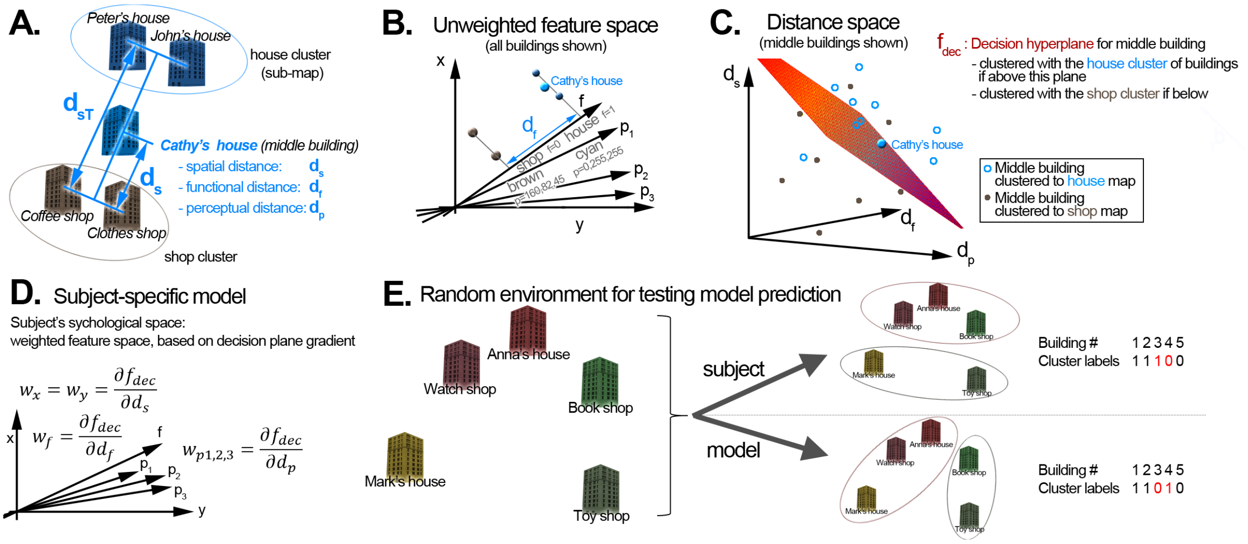

A.
Peter's house
John's house
house cluster
(sub-map)
d_{sT}
Cathy's house (middle building)
- spatial distance: d_s
- functional distance: d_f
- perceptual distance: d_p
d_s
Coffee shop
Clothes shop
shop cluster
B. Unweighted feature space
(all buildings shown)
x
y
Cathy's house
d_f
f
shop f=0
house f=1
p_1
p_2
p_3
brown p=160,82,45
cyan p=0,255,255
C. Distance space
(middle buildings shown)
f_{dec} : Decision hyperplane for middle building
- clustered with the house cluster of buildings if above this plane
- clustered with the shop cluster if below
d_s
d_f
d_p
Cathy's house
○ Middle building clustered to house map
● Middle building clustered to shop map
D. Subject-specific model
Subject's sychological space:
weighted feature space, based on decision plane gradient
$w_x = w_y = \frac{\partial f_{dec}}{\partial d_s}$
$w_f = \frac{\partial f_{dec}}{\partial d_f}$
$w_{p1,2,3} = \frac{\partial f_{dec}}{\partial d_p}$
x
y
f
p_1
p_2
p_3
E. Random environment for testing model prediction
Anna's house
Watch shop
Book shop
Mark's house
Toy shop
subject
model
Building # 1 2 3 4 5
Cluster labels 1 1 1 0 0
Building # 1 2 3 4 5
Cluster labels 1 1 0 1 0

Figure 8: The decision hyperplane method for inferring feature importances and generating environments in Experiment 2. A: General layout of training trials, which consisted of two groups of two buildings (with equal colour and function), and a middle building, the parameters of which could be varied (distance, similarity in colour and in function to the shop group). B: Feature space representation - each building can be represented as a single point in a space spanned by the features (position, colour, function). C: Distance space representation - middle buildings can be represented in terms of their distance to the shop group along each feature. According to the clustering hypothesis, there has to be a 'decision hyperplane' calculable from these middle buildings, such that those below the plane (i.e. those closer to the shop group) are most likely clustered with the shop group, and those above the plane (i.e. farther away from the shop group) are most likely clustered with the house group. D. Subject-specific models consist of a weighted feature space and a clustering algorithm. The weights of the feature space can be calculated from the decision boundary - the importance of each feature is proportional to the derivative (slope) of the decision boundary by that feature. E. Randomly generated testing environments, and comparison procedure. Subjects impose a grouping even on random, unstructured environments, as shown by previous research (McNamara et al., 1989). The clustering model also produces a grouping, based on the learned subject-specific model and the clustering algorithm. Subsequently, cluster labels are compared (and, in this example, found to be incorrect).

in which the model parameters $W = (w_s, w_f, w_p, ...)$ control the slope of the decision hyperplane, and thus represent participants' feature importances in this model. They were used to construct the participant's 'psychological space', i.e. a feature space weighted by these parameters (as illustrated in Figure 1B), which lead to attenuated differences along features unimportant to the participant.

Subsequently, we used clustering in this weighted feature space for prediction. We employed the DP-GMM (Dirichlet Process Gaussian Mixture Model), from the family of Bayesian nonparametric models, for clustering (see Supplementary Information for the mathematical formulation and (Gershman & Blei, 2012) for a tutorial review). Bayesian nonparametric models were successfully employed in categorization models (Sanborn et al., 2006) and shown to be psychologically plausible, unifying previously proposed models of category learning (Griffiths et al., 2007) and accounting for several cognitive mechanisms including category learning and causal learning (Tenenbaum et al., 2011), transfer learning (Canini et al., 2010), and human semi-supervised learning (Gibson et al., 2013). Given that such models give a good account of how humans acquire novel concepts (subsuming prototype, exemplar, and rational models of category learning, among others), and given that they can be seen as probabilistic clustering models, we hypothesized that they might also account for sub-map learning.

DP-GMMs are extensions of Gaussian Mixture Models (GMMs) for an unlimited number of clusters. GMMs are statistical models which aim to partition a set of data points in some space into a number of clusters C by fitting C Gaussian probability distributions to the data, i.e. adjusting the parameters of these C Gaussians such that the probability that the data was drawn from these distributions is maximized. DP-GMMs have the same aim, but also allow inferring the number of distributions (and thus the number of clusters C), not just their parameters. In this lies their key advantage compared to most other clustering models: they can be used without prior knowledge of the correct number of clusters (and they can expand by adding new points either to the most likely existing cluster, or to a novel cluster, when observing new data). This process of assigning new data points to clusters by calculating probabilities from distributions optimally fitted to previous data has a lot in common with the basic problem of categorization, which is to identify the category of a new object based on its observed properties and previously observed objects, which is why Bayesian nonparametric models are similar to (in fact, if parametrized accordingly, mathematically equivalent to) multiple psychological models of category learning proposed in the past (Griffiths et al., 2007).

The final sub-map membership predictions were generated by performing clustering, using a DP-GMM[15], in the weighted feature space learned from the subject. These predictions were evaluated by calculating prediction accuracies and Rand indices (Rand, 1971). The former is simply the ratio of perfectly predicted sub-map structures to all subject structures - however, this strict accuracy metric penalizes 'near misses' equally to completely wrong structure predictions (e.g. if seven building sub-map memberships are correct, but a single one incorrect, the entire prediction is counted as incorrect; just like completely wrong structures). Average Rand indices are reported as more fair metrics which provide a continuum between flawlessly correct ($R = 1$) and completely incorrect ($R = 0$) predictions. The Rand index is a measure of the amount of correctly assigned pairs among all pairs, and is defined as $R = (s + d)/\binom{B}{2}$, where B is the total number of buildings on a map structure, s is the number of building pairs on the same sub-map both in the predicted and actual map structure, and d the pairs on different sub-maps both in prediction and in subject data.

3.4.2. Participants

Participants were students at the University of Manchester (compensated by vouchers). Subjects who did not produce sketch maps significantly better than random chance in at least 50% of all training trials were excluded, leaving 12 subjects whose data was analysed. Participants were told they need prior experience with either virtual realities or three-dimensional computer games. These participants were recruited and tested at the University of Manchester (instead of online) primarily because the setup required a modern PC equipped with a graphics card to run the experiment smoothly. Further reasons were the requirement of prior 3D gaming experience (difficult to verify online), and the need to ensure that the setup was equivalent across subjects (e.g. screens were of the same size and quality, all subjects used a mouse and not a touchpad, etc.).

3.4.3. Procedure

After giving their consent and reading instructions, participants completed 20 trials - 15 'training' trials which were used for training the model, and 5 'testing' trials which were used for verifying the predictions of the computational model. In total, the experiment took about 1.5 hours on average. Each trial was set in a unique environment consisting of a horizontal ground plane, featureless sky, and 5 buildings. All buildings used the same 3D model and thus had equal measurements, but could vary in colour, in function (being labelled as either shops or houses) and in distance; and could have different labels (e.g. coffee shop, John's house).

Both trial types followed the same sequence. First, participants could freely explore the environment, and were asked to memorize the positions and names of all buildings in it. In this memorization phase, they were also asked to deliver a package from one of the shops to one of the houses. This task served the dual purpose of forcing subjects to do a minimum amount of exploration, and, additionally, to make the functional distinction between shops and houses more meaningful. After the memorization phase and the delivery task, the environment vanished, and participants' spatial memory was tested, by asking them for 1) a sketch map, produced by dragging and dropping labelled squares into their correct places, and 2) seven recall sequences, 5 cued, and 2 uncued.

The first 15 'training trials' each contained two distant groups of two buildings in close proximity, and a 'middle' building somewhere between these two groups. Both buildings in each group always had the same

[15] With variational inference to infer the most likely cluster memberships and parameters. We have used the *bnpy* Python library for inference (Hughes & Sudderth, 2013)

colour and function, and there was always one group containing two shops and a second group containing two houses. The middle building was intended to be represented together with one or the other group by subjects, depending on its distance and similarity to the groups. In the first 7 of these trials, the colours, functions, and distances of the groups and the middle building were generated randomly, ensuring only within-group consistency of colour and function and that buildings within groups were closer than the distance between the groups, such that they unambiguously formed clusters.

After the 7th[16] training trial and all subsequent training trials, a 'decision hyperplane' was calculated using logistic regression, which separated all middle buildings into two groups, those belonging to the shop cluster, and those belonging to the house cluster. This decision hyperplane facilitated the generation of the remaining 8 training trial environments. For each trial after the 7th, the two groups were again generated randomly, but the middle building was parametrized such that the uncertainty regarding subjects' feature importances was minimized. To achieve this, the parameters of the new middle building were drawn from the region of the currently calculated decision hyperplane, since this is the region in which the model is least certain as to where buildings should be assigned[17]. Formally, this is equivalent to active learning (Settles, 2010) with uncertainty sampling (Lewis & Gale, 1994) in machine learning. Each of these remaining 8 training trials maximally reduced the model uncertainty regarding feature importances.

Finally, participants completed 5 'testing' trials, going through the same procedure of memorization, delivery task, and producing a sketch map and recall sequences. These testing trials were generated completely randomly, without any restrictions on building parameters, not even the restriction of there needing to be clusters defined in any way along any of the features. They were used to test the predictions of the computational model.

3.4.4. Results

We included all features described in Section 3.3 in the following analysis, except for the four geospatial features not relevant in our simple virtual reality environment (path distance, natural boundaries, number of intersecting streets, whether they could be easily crossed). For the correlations of these features with co-representation probabilities (Figure 6), as well as across- and within-subject variances of these correlations (Figure 7), see Section 3.3.

Above, we have introduced a method to infer participants' feature importances for clustering, based on the inference of a decision hyperplane describing at which point in feature space subjects stop assigning a middle building to one sub-map and start assigning it to another. With this method, we have both components of a predictive model of cognitive map structure: 1) subjects' psychological spaces, spanned by a set of features and feature importances, as inferred by the decision hyperplane approach, and 2) a clustering algorithm. We chose DP-GMM as the clustering algorithm, given its substantial advantage of being able to infer the number of sub-maps automatically, and motivated by its success in other psychological models.

Figure 9 shows the results of this predictive model on all participant cognitive map structures (20 per subject; 15 training maps used to infer feature importances, and 5 testing maps used to verify model predictions). Prediction can be incorrect on training trials, because feature importances are being inferred using the decision hyperplane approach without taking into account the clustering algorithm and its idiosyncrasies (see red cells of the first 3 rows). After inference of feature importances and running the clustering model within this feature space, 73.5% of the training map structures could be predicted.

The interesting part of Figure 9 is the bottom row of each sub-plot, which contains the advance predictions of the model on randomly generated environments it was not trained on and not confronted with prior to

[16] Due to the noise inherent in the inference of building map memberships, using active learning from the start yielded bad results, the model hypothesizing a highly sub-optimal decision hyperplane it could not recover from using the limited number of subsequent data points. A few randomly initialized trials were used at the start to avoid this and to allow the inference of an approximate decision hyperplane before starting the active learning process. Empirical experimentation using artificially generated maps, and an amount of outliers comparable to subjects, suggested 7 random and 8 uncertainty sampling trials, when given 15 datapoints (from the 15 trials).

[17] As the region of least certainty, or greatest uncertainty, comprises the points with a classification probability of 0.5 to either class, these points can be defined as: $D_{LC} = \underset{D}{\operatorname{argmin}} |0.5 - P(S|D)|$. From this and eq. (1), it follows that $W^T D_{LC} = 0$, i.e. that points of least certainty lie on the hyperplane described by W, confirming the informal argument in the text above.

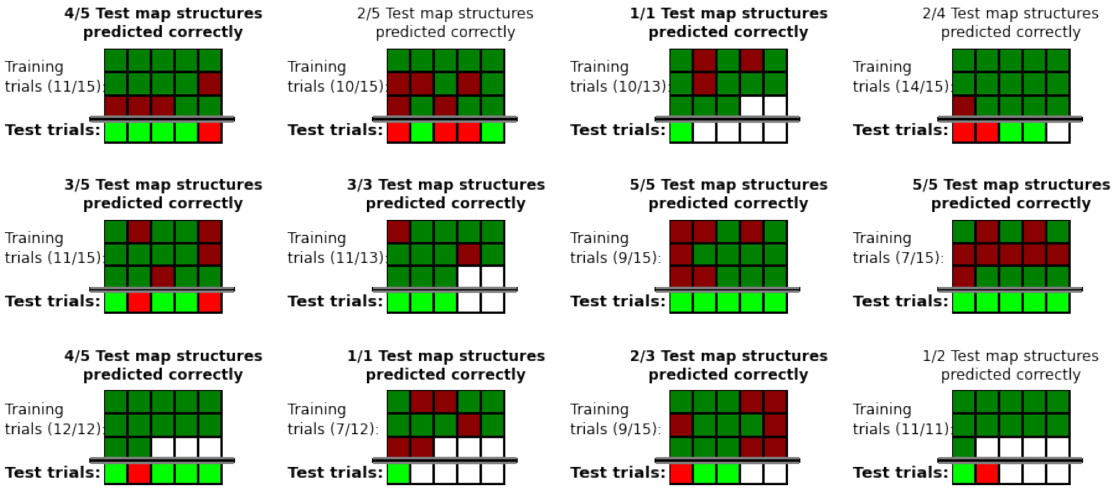


Figure 9: Results of a predictive clustering model using subjects' feature importances, learned using the decision hyperplane approach. Each sub-plot reports all prediction results for one subject, using green cells for correct predictions, red cells for incorrect predictions (one or more buildings grouped to the wrong sub-map), and white cells for subject maps either not better than random chance or without apparent structure. Top 3 rows in each subplot show results on the training trials (dark colours), and the 4th, bottom row shows the prediction accuracies on the test trials (bright colours). On average, 75% of all test map structures could be predicted correctly (green cells). For comparison, the probability of prediction by random chance is 0.4% for two sub-map and 3.1% for one sub-map structures.

making the prediction. On average, **75.0% of all test map structures could be predicted correctly in advance** using the decision hyperplane method and DP-GMM for clustering; and the majority of map structures could be predicted for all subjects except for one.

Note that this is a strict accuracy metric - if the model predicts four out of five building sub-map memberships correctly, but a single one incorrectly, the entire prediction is counted as incorrect. The Rand index (Rand, 1971) is a more comprehensive metric, providing a number between 1 (flawless clustering) and 0 (all cluster memberships incorrect). The **average Rand index of predicted vs. actual test map structures was 0.83** in this experiment, meaning that for 83% of the pairs of buildings, it could be correctly predicted whether or not they belong to the same sub-map in participants' spatial memory (according to their recall sequences).

If using the same DP-GMM model with feature importances inferred from co-representation correlations instead, the prediction accuracy drops to 59.1% on the testing maps, with an average test-map Rand index of 0.75, indicating that the decision hyperplane approach is better suited to uncovering feature importances than just using correlations.

Since each environment contained five buildings, there could be up to two sub-maps, and the clustering process could be framed as assigning one of three values to each building - member of sub-map #1, or of sub-map #2, or a single-building cluster (sub-maps with only a single building were excluded from participant data, for reasons explained in Section 2; however, if the model produced single-building clusters, these were not excluded from the model predictions, but instead counted as mistakes). Thus, the baseline probability of randomly coming up with the correct clustering is, on average, $(1/3)^5 = 0.4\%$ for map structures with two sub-maps, and $(1/2)^5 = 3.1\%$ for structures with one sub-map of unknown size. In this experiment, 14 subject test map structures contained two sub-maps, and 30 structures one sub-map.

3.4.5. Discussion

The observation that a large majority of subject map structures can be predicted in advance using a clustering model, together with an appropriately scaled feature space, provides further support for the clustering hypothesis. The improvement of prediction accuracy from 59.1% to 75.0% (and Rand index from 0.75 to 0.83) when using the decision hyperplane approach to infer feature importances, instead of just using co-representation correlations, suggests that this approach is more suitable to uncover the psychological spaces in which the clustering takes place.

However, the present approach has several shortcomings. First, it is only applicable to controlled environments - thus, investigating participants' past long-term memory structures requires different methods (see next section). Second, the fact that calculating feature weights from a decision hyperplane does not take into account the actual model generating the predictions (in this case, the DP-GMM). Finally, the approach assumes linearity, i.e. that the surface separating buildings co-represented with one or the other sub-map is a linear hyperplane (as opposed to a non-linear surface). These shortcomings are reflected in the sub-optimal performance of the model on the training trials in Figure 9. Although a model should be able to fit its training data well, the performance on training trials (73.5) and testing trials (75%) is not statistically significantly different.

Thus, it is likely that more powerful models to learn subjects' feature spaces are needed. The next section introduces two such approaches addressing these shortcomings, one learning the optimal feature weights for the employed clustering model using global optimization, and the second lifting the linearity assumption. Both of them have the additional advantage that they do not require controlled environments.

3.5. Experiment 3 - Predictability of cognitive map structure in the real world

In this experiment, real-world buildings well known to participants were used (similarly to Exp. 1). Apart from providing additional evidence for the clustering hypothesis by showing that cognitive map structures in real-world environments can be predicted using a clustering model, this section also introduces and validates two generally applicable ways of learning subject-specific models.

3.5.1. Computational methods

Unlike in the previous section, where participants' feature importances were inferred using the decision hyperplane method - which requires controlled environments (Section 3.4), we use two generally applicable methods in this section (see Figure 10):

1. Global optimization (Jones et al., 1993)[18] - among all possible feature weights (between 0 and 1), select the features and weights best explaining the groupings of the 'training' structures obtained from the participants (a part of each participant's data was used for training, and the rest for 'testing', i.e. prediction verification). Use clustering in this weighted feature space for prediction.
2. GDA (Gaussian Discriminant Analysis) (Bensmail & Celeux, 1996) - using the set of all training building pairs, learn a probabilistic (Gaussian-based) model capable of calculating the probability of whether any given pair of buildings are co-represented on the same sub-map, given the distances of this pair along various features. Use this probabilistic model as a distance metric[19] (such that building pairs which are likely to be on the same representation are close, and those which are not are distant, under this metric). Predict subject map structures by clustering under this learned, subject-specific metric. See Supplementary Information for the mathematical formulation.

The first of these two, as well as the hyperplane approach, are both linear methods, whereas the latter method (GDA) allows non-linear solutions. Linear methods project data into psychological space by linearly

[18] We used the locally biased variant (Gablonsky & Kelley, 2001) of DIRECT (DIviding RECTangles) (Jones et al., 1993), a global, deterministic, derivative-free optimization method based on Lipschitzian optimization, which can handle the kinds of non-linear and non-convex functions which clustering accuracy inevitably entails. DIRECT finds global optima by systematically dividing the feature space into smaller and smaller hyperrectangles, returning the one yielding the best results upon convergence.

[19] For a pair of buildings represented by feature vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, given their absolute difference $\Delta\mathbf{x} = |\mathbf{x}_1 - \mathbf{x}_2|$ as well as a trained GDA model which is able to calculate the probability $p(c = 1|\Delta\mathbf{x})$ that these buildings are co-represented on the same sub-map, based on appropriately fitted Gaussian distributions (Bensmail & Celeux, 1996), we simply define the metric as $d_{Metric}(\mathbf{x}_1, \mathbf{x}_2) = 1 - p(c = 1|\Delta\mathbf{x})$, where the probability of co-representation is derived using Bayes rule, $p(c = 1|\Delta\mathbf{x}) \propto p(\Delta\mathbf{x}|c = 1)p(c = 1)$, and the generative densities are modelled using multivariate Normal distributions $p(\Delta\mathbf{x}|c = 1; \mu_i, \Sigma_i) = (2\pi)^{-\frac{D}{2}} |\mathbf{\Sigma}_i|^{-\frac{1}{2}} e^{-\frac{1}{2}(\Delta\mathbf{x} - \boldsymbol{\mu}_i)^\intercal \mathbf{\Sigma}_i^{-1} (\Delta\mathbf{x} - \boldsymbol{\mu}_i)}$ (see Supplementary Information for details).

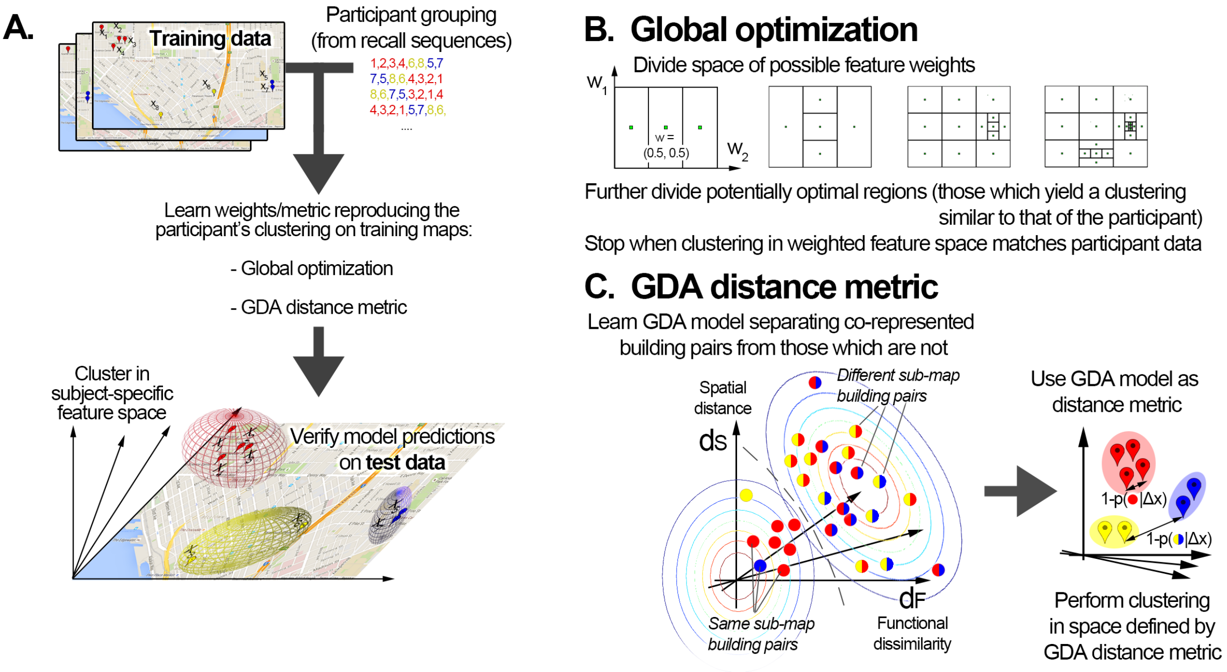

A.
Training data
Participant grouping
(from recall sequences)
1,2,3,4,6,8,5,7
7,5,8,6,4,3,2,1
8,6,7,5,3,2,1,4
4,3,2,1,5,7,8,6,
....
Learn weights/metric reproducing the participant's clustering on training maps:
- Global optimization
- GDA distance metric
Cluster in subject-specific feature space
Verify model predictions on test data
B. Global optimization
Divide space of possible feature weights
w1
w = (0.5, 0.5)
w2
Further divide potentially optimal regions (those which yield a clustering similar to that of the participant)
Stop when clustering in weighted feature space matches participant data
C. GDA distance metric
Learn GDA model separating co-represented building pairs from those which are not
Spatial distance
ds
Different sub-map building pairs
dF
Functional dissimilarity
Same sub-map building pairs
Use GDA model as distance metric
1-p(●|Δx)
1-p(●|Δx)
Perform clustering in space defined by GDA distance metric

Figure 10: Learning subject-specific models for predicting cognitive map structure. A: General modelling procedure. Several sets of buildings and their groupings are obtained from different environments using the recall sequence paradigm, and these are split into two parts, training data and test data. A subject-specific model is learnt such that clustering under this model (e.g. in an appropriately weighted feature space) reproduces the training data as well as possible. Finally, model predictions are generated by clustering test buildings not seen at training time, under the learned models, and these predictions are compared to the actual participant map structures (groupings) in the test data. B: A weighted feature space (modelling participants' 'psychological space') can be obtained by searching for the optimal weights using global optimization. This method keeps dividing the space of possible feature weights into thirds, and further divides potentially optimal regions (those which correspond to feature weights under which clustering yields a grouping close to participant's map structure), until the weights best matching participant training data are found. C: A metric space modelling participants' psychological space can also be defined by a non-linear metric instead of linear feature weights. Such a metric can be learned by fitting a GDA model to separate building pairs that belong to the same sub-map from those that do not. In order to make predictions, building representations are projected into a space where their distances are dictated by this GDA model (such that they are close if they are likely to belong to the same sub-map); and clustering is performed in this space.

weighting the features, and try to find weights such that buildings co-represented on the same sub-maps are closer to each other in this weighted feature space than other buildings (note that the problem of projecting the data into a subject-specific feature space with some learned weights is equivalent to finding a distance metric with those feature weights[20]). In addition to these linear methods, we wanted to test a more powerful method that can capture non-linearities as well as interactions between the features (e.g. situations where the importance of one feature depends on the magnitude of another). We implemented a novel method, instead of using existing metric learning approaches, see (Yang & Jin, 2006) for a review, for the following reasons. First, our method can naturally incorporate the hypothesis that same sub-map building pair differences should be small, and thus located close to the origin, and should be separable from different-map building pair differences (these two distributions of pair differences can be naturally modelled using Gaussian distributions) - see Figure 10C. Second, our data violates some of the assumptions of existing methods[21]. Third, most existing machine learning solutions - as well as MDS, used in cognitive

[20] This equivalence is easy to see from rewriting the equation of the Mahalanobis distance metric, $d_M(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{(\mathbf{x}_2 - \mathbf{x}_1)^\intercal W (\mathbf{x}_2 - \mathbf{x}_1)}$, to the following form: $d_M^2(\mathbf{x}_1, \mathbf{x}_2) = ||A(\mathbf{x}_2 - \mathbf{x}_1)||$, where $W = A^\intercal A$, and A is a projection map that can transform data x into weighted feature space $y = Ax$

[21] Metric learning is concerned with finding a distance metric - such as linear, Mahalanobis distance metrics, and their associated parameters, e.g. (Xing et al., 2002), or non-linear metrics by projecting the data into kernel space using e.g.

psychology to model similarities as distances (Shepard, 1957) - need to embed both training map and test map buildings into the same space for model training and testing. This is not possible in our case, because 1) for the features of functional and perceptual similarity, the pairwise similarities across environments are unknown (since subjects only indicate these within each map, not across maps), and 2) spatial distances might not be comparable across cities or countries (whether two buildings belong to the same representation strongly depends on their geographical distance; but this dependence likely becomes weak or non-existent if they are very far apart).

After a subject-specific model has been learned, sub-map memberships can be predicted by performing clustering based on this model (i.e. within the feature space / under the metric learned from the subject). Just like in the previous section, we used the DP-GMM clustering algorithm for this purpose.

Before reporting prediction results, we should point out that there are theoretical as well as practical limits on the predictability of cognitive map structures. Section 4.3 discusses these in more detail and suggests some solutions. Here, we shall focus on the main issue concerning data analysis, namely detecting and removing outliers caused by distractions or lapses of attention. If a set of buildings that are actually co-represented on a sub-map in a subjects' spatial memory is recalled together most of the time, but the subject is distracted during one of the recall sequences, and recalls a different (not co-represented) building instead, the subsequently extracted structure will be incorrect (since tree analysis requires items to occur together in *every* recall sequence in order to identify a sub-map). Even a single distraction during the 7 or 10 (in Experiment 3 A or B) recall sequences per trial can yield substantially different structures (see example in Figure 13 in Section 4.3, in which a distraction causes a drop of 0.6 in the Rand index to the correct structure).

The jackknifing procedure we use to eliminate outliers was suggested by the authors pioneering the recall order paradigm (Hirtle & Jonides, 1985; McNamara et al., 1989) to mitigate this issue, but relies on statistical significance testing to identify those outliers, and thus frequently fails to do so due to the small number of recall sequences collected in our experiments (a necessary limit arising from the need to collect multiple different map structures for training and testing a predictive model - subjects already took up to 3.5 hours for these experiments even with this small number of sequences).

It is possible to estimate the effectiveness of jackknifing in our data - and the percentage of incorrectly inferred and thus unpredictable map structures resulting from it (see Figure 11). To do this, we simulated distractions by randomly swapping two items in one of the sequences in each trial. This is a reasonable model of distractions, since the only way subjects can make mistakes is by changing the order of their input (they are forced to repeat the trial if they omit or incorrectly recall an item).

The *number* of simulated distractions (frequency of swapped items) makes no difference to the estimated *percentage* of outliers that are not caught and excluded by jackknifing. We used one distraction per trial (however, the following results stayed the same with 0.5 or 2 distractions per trial). For the 5 buildings maps (and 7 recall sequences), and averaging over 100 runs, each with a single random non-cue lapse for all subjects, simulated distractions cause changes in map structure (relevant outliers) in $\mu_n = 65.4\%$, $\sigma_n = 3.7\%$, and within these, outlier removal is effective in $\mu_e = 59.4\%$, $\sigma_e = 5.0\%$. The situation is somewhat better on the 8 building maps, due to the larger numbers of sequences collected and thus higher statistical power - here, outlier removal is effective in $\mu_e = 56.0\%$, $\sigma_e = 8.1\%$ of the cases (and necessary only in $\mu_n = 33.2\%$, $\sigma_n = 6.0\%$). This leaves on average $\mu_u = 26.6\%$ ($\sigma_u = \sqrt{\sigma_e^2 * \sigma_n^2 + \mu_e^2 * \sigma_n^2 + \mu_n^2 * \sigma_e^2} = 3.9\%$) of disruptive simulated lapses of attention for condition A, and $\mu_u = 15.0\%$ ($\sigma_u = 4.3\%$) for condition B, which cannot be mitigated by jackknifing.

If we assume this uniform random swapping to be a reasonable approximation of subject distractions, this would mean that apart from the approximately $o = 9.5\%$ of sequences which were successfully removed

a Radial Basis Function (RBF) kernel Φ in a distance function $d_{RBF}(x_1, x_2) = \sqrt{(\Phi(x_1) - \Phi(x_2))^\intercal (\Phi(x_1) - \Phi(x_2))}$, e.g. Baghshah & Shouraki (2010); Chitta et al. (2011). However, the former makes the assumption of *linear separability*, and the latter require *variances to be isotropic*, i.e. to not differ much across features (since the RBF kernel uses a diagonal covariance matrix, it cannot fit non-isotropic data well - see Ong et al. (2005)). As both of these assumptions are occasionally violated in our subject data, these metric learning approaches are not applicable. In contrast to existing metric learning, our proposed approach learns a probabilistic model in the space of pairwise differences (instead of learning from scalar distance values, it learns from difference vectors), and thus can fit non-isotropic and non-linear data.

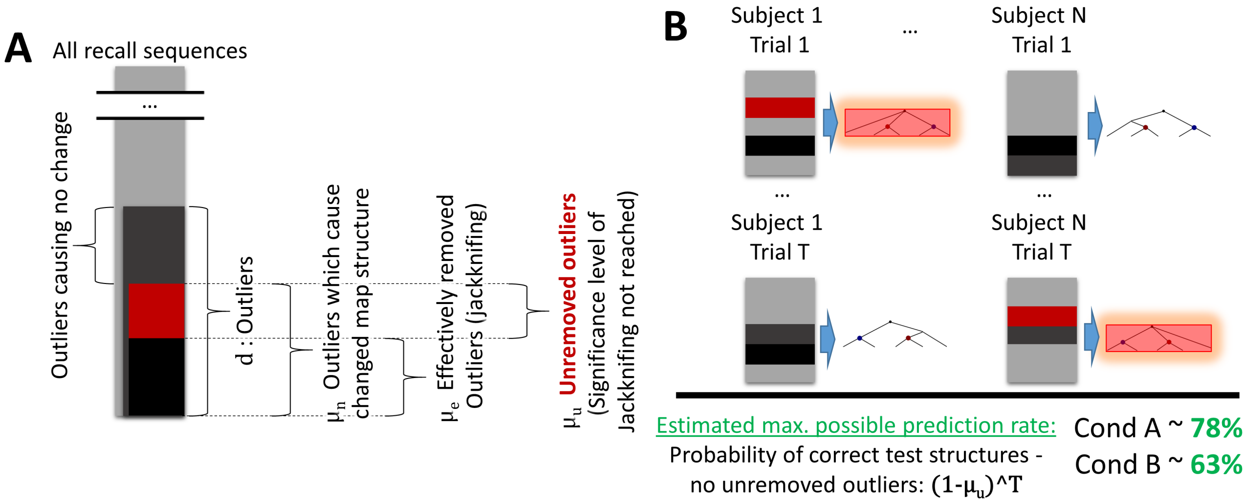

A
All recall sequences
...
Outliers causing no change
d : Outliers
μ_n Outliers which cause changed map structure
μ_e Effectively removed Outliers (jackknifing)
μ_u Unremoved outliers (Significance level of Jackknifing not reached)
B
Subject 1
Trial 1
...
Subject N
Trial 1
...
...
Subject 1
Trial T
Subject N
Trial T
Estimated max. possible prediction rate:
Probability of correct test structures - no unremoved outliers: $(1-\mu_u)$^T
Cond A ~ 78%
Cond B ~ 63%

Figure 11: Estimated maximum possible prediction rate using the data in Experiment 3. A: Assuming that distractions / lapses of attention manifest as randomly swapped items in recall sequences (and cause changes in the inferred tree structures), a substantial number of them cannot be detected using the outlier detection procedure (jackknifing) proposed in the seminal work on hierarchical cognitive maps and employed in this paper. B: Undetected outliers in recall sequences cause a number of inferred map structures to be incorrect. This results in a percentage of map structures not predictable even by good models.

as outliers using jackknifing (and thus part of the effective 59.4% or 56% for cond. A and B), there would be an expected additional $\mu_o = 6.5\%$ ($\sigma_o = 0.1\%$) of sequences for condition A, and expected $\mu_o = 2.5\%$ ($\sigma_o = 0.7\%$) for condition B, which would likely be outliers causing structure changes which have not been removed by jackknifing because of the lack of statistical significance. It follows that the expected probability of extracting correct map structures under these assumptions - and thus the maximum possible prediction rate - is around $(1 - 0.065)^7 \simeq 63\%$ for condition A (since there are 7 sequences per trial), and around $(1 - 0.025)^{10} \simeq 78\%$ for condition B (since there are 10 sequences per trial).

To summarize, the observation that not all simulated distractions (outliers) can be identified and omitted by the jackknifing procedure strongly suggests that the data collected from human subjects also contains outliers not caught by jackknifing. Thus, these outliers prevent perfect prediction of subject map structures. Figure 11 summarizes this reasoning and the maximum possible prediction rates estimated based on it for both conditions.

3.5.2. Participants

Data from 71 participants was analysed in this section, 54 in Experiment 3A (asked for 5 environments with 5 buildings each), and 19 in Exp 3B (asked for 3 environments with 8 buildings). Subjects unable to produce at least two sketch maps significantly better than random chance (see Section 2.3), with structure apparent from their recall sequences for at least two maps, were excluded, as at least two map structures were required to have both a training and testing map. Participants were recruited, consented, and compensated through the Amazon Mechanical Turk online survey system, and were required to have at least 95% approval rating on previous jobs to ensure higher data quality.

3.5.3. Procedure

The procedure was similar to the one used in Experiment 1. This experiment was also conducted on a website participants could access through MTurk after giving their consent. Unlike 1, this experiment consisted of multiple trials (5 in condition A, 3 in condition B), each trial following an equivalent procedure but asking for a completely different set of buildings, possibly in a different city. Subjects took between one and 3.5 hours to complete this repeated trial experiment (this includes possible breaks, since the experiment was performed online in participants' homes, unsupervised, and the experiment was not timed).

In the first questions of each trial, subjects were asked to pick a number of buildings they know well - 5 in condition A, and 8 in condition B (thus, in total, 25 buildings had to be recalled for the 5 trials of condition A, and 24 for the 3 trials of condition B). Thus, well-memorized long-term memories of real-world environments were tested instead of novel stimuli in virtual reality. Subjects were instructed to make sure that they know where in the city these buildings are located, how to walk from any one building to any of the others, what each building looks like, and what purpose it serves.

The subsequent questions of each trial required subjects to produce a sketch map, and to perform a recall test consisting of 7 recall sequences in condition A, and 10 in condition B (in both cases, as many cued sequences as there were buildings on the maps, and two additional uncued sequences). Subjects followed the same instructions as in Experiment 1 ; the crucial difference being that instead of presenting cues verbally by writing out the name of the cue building, cues were presented visually (cue modality was changed to mitigate the strong effects of phonetic and morphological similarity in the prior experiments, presumably due to articulatory rehearsal strategies). Participants were shown building positions on their own sketch maps prior to each recall sequence question, excluding the labels - only the uniform gray squares symbolizing the buildings were shown. For each cued recall question, the cue (starting building) was indicated by highlighting the cue building in green colour and with a thick border.

In the final question, subjects were asked to judge the similarities of all pairs of buildings, i.e. $\binom{5}{2} = 10$ pairs in condition A and $\binom{8}{2} = 28$ pairs in condition B, as well as a control pair of one of the buildings to itself, both in terms of visual similarity, and similarity of purpose/function (using 1-10 rating scales as before).

3.5.4. Results

Figure 12 shows prediction accuracies (the ratio of perfectly predicted map structures to all subject map structures) using DP-GMM clustering and GDA subject-specific model learning. Using the best possible set of features shown to the model[22], **68.6% of the 185 subject map structures with 5 buildings of Experiment 3A** (with up to two sub-maps per structure), and **79.2% of the 48 subject map structures with 8 buildings of Experiment 3B** (with up to four sub-maps per structure) **can be predicted accurately**, such that every single predicted sub-map membership is correct for these percentages of test maps. **Average Rand indices for these models are 0.87 for condition A and 0.95 for condition B**, which means that even the structures which are imperfectly predicted, causing a lower than optimal prediction accuracy, are highly similar to the correct structures (co-represented building pairs are predicted correctly in 87% in condition A and 95% in B). Note that the prediction accuracy of the best model is statistically indistinguishable from the estimated maximum possible prediction rate (calculated above based on simulating distractions by random swapping). This suggests that the proposed novel GDA-based method does well at learning subject feature spaces, and that the subsequent clustering model, based on a previously proposed Bayesian model of category learning, can infer the sub-map memberships and numbers accurately.

Figure 12 also shows the numbers of sub-maps contained in participants' structures. In general, the prediction task can be seen as assigning one of $K + 1$ values to each building, where K is the maximal number of possible sub-maps (single-building clusters are also possible, hence the increment by one). Thus, the baseline probability of randomly coming up with the correct clustering is, for condition A, $(1/3)^5 = 0.4\%$ for map structures with two sub-maps, and $(1/2)^5 = 3.1\%$ for structures with one sub-map. For condition B, this baseline expected random clustering accuracies are several orders of magnitude lower ($2.5 * 10^{-4}\%, 1.5 * 10^{-3}\%, 1.5 * 10^{-2}\%$ and 0.3% respectively for $K = 4, 3, 2$ and 1).

The model accuracies when successively removing particular features (bars from left to right in Figure 12) provide an additional measure for how important these features were, aggregated over all subjects, and measuring importance in a causal fashion, since this is a predictive model. The most important features were those which caused the greatest drops in accuracy upon their removal. In condition A, two features

[22] Estimated from the training data, using a greedy search approach - starting with a single feature (Euclidean distance) and then iteratively adding the feature which brings the clustering prediction closest to participants' actual groupings; repeated until either all features are included or the clustering prediction accuracy stops increasing.

Condition	All features, subject-specific GDA model	A-priori features subject-specific GDA model	No subject-specific model
Condition 3A	**79.2 % (RI=0.94)**	70.8% (RI=0.88)	41.7% (RI=0.76)
Condition 3B	**68.6% (RI=0.89)**	63.4% (RI=0.83)	60.2% (RI=0.78)

Table 2: Prediction accuracies (and Rand indices) in Experiment 3, for all features and subject-specific GDA+DP-GMM model (second column), for features known a-priori, without having to ask subjects to rate similarities or draw sketch maps (third column), and finally using a subject-general model, without learning subject-specific feature weights. Rows: Condition 3A (19 subjects, 48 map structures from as many distinct environments, 112 sub-maps), and condition B (54 subjects, 185 map structures from as many distinct environments, 310 sub-maps).

are significantly more important than the rest - sketch map distance and the product of path distance and visual similarity -, whereas the importances are similar in condition B, with a slightly larger accuracy drop caused if omitting sketch map distance. In both conditions, about 2 out of 5 map structures can still be predicted when using solely Euclidean distance.

The strong influence of sketch map distances raises an additional question regarding predictability of cognitive map structures - is advance prediction possible without asking the subject anything (other than a list of buildings he knows)? To investigate this question, we have run the predictive model on data from which visual similarities and sketch map distances were removed, i.e. solely on data which can be derived from the list of subjects' buildings (see Section 3.3 for geospatial data sources). Subjects' functional similarities were also removed from this data, and replaced by an objectively calculated measure of functional relatedness. Specifically, we used the Jaccard similarity metric on lists of building types from Google Places API [23]. The objective functional similarity metric thus obtained does reflect subjects' own judgements - the correlation between them is $r = 0.66$ - but is somewhat different, since it does not reflect subject idiosyncrasies, and is also free of noise or biases.

Using GDA for subject-specific model inference, and using these features which are all known a priori - derivable from the subject building lists and public geospatial databases -, 75% of map structures can be predicted in advance for condition A (Rand index: 0.91), and 68.8% in condition B (Rand index: 0.88).

Finally, we have attempted to predict subjects' cognitive map structures without learning subject-specific models at all, by trying to infer a psychological space common to all subjects, and clustering within this space. Inferring someone's spatial representation structure without knowing anything about them would have great advantages for robotics applications and geographical planning and map design, among other fields (see Section 4.1). The resulting prediction accuracies (and Rand indices) for condition A and B were 41.7% and 60.2% (and $RI = 0.76$ and 0.78) respectively. In accordance with the results in Section 3.3, the model performs significantly worse when not allowed to learn subject-specific feature spaces. However, even these impoverished models can predict whether or not two buildings are co-represented on the same sub-map in more than 3 out of 4 cases.

3.5.5. Discussion

The model prediction accuracies reported above are close to the estimated maximum possible prediction rates from noisy map structures (based on simulating participant distractions using random swapping), calculated at the beginning of this Section: 62.5% for condition A, and 78.0% for condition B. This shows that the model accounts well for this noisy data, despite not being able to predict 100% of subject map structures.

[23] Places API can return a list of known building types when queried - see `https://developers.google.com/places/supported_types` for a list. Usually buildings have several applicable types, ranging from specific to general, e.g. 'meal takeaway', 'restaurant' and 'food' for McDonalds. The Jaccard index (JI), defined as the ratio of the size of the intersection to the size of the union of two sets, measures how many items in these type lists match between two buildings, as a proxy for their functional similarity. For example, $JI = 0.5$ between the McDonalds example and a building with types 'bakery', 'restaurant' and 'food'. The type 'establishment' was present for almost all buildings and was thus excluded from the computation of JIs, being uninformative.

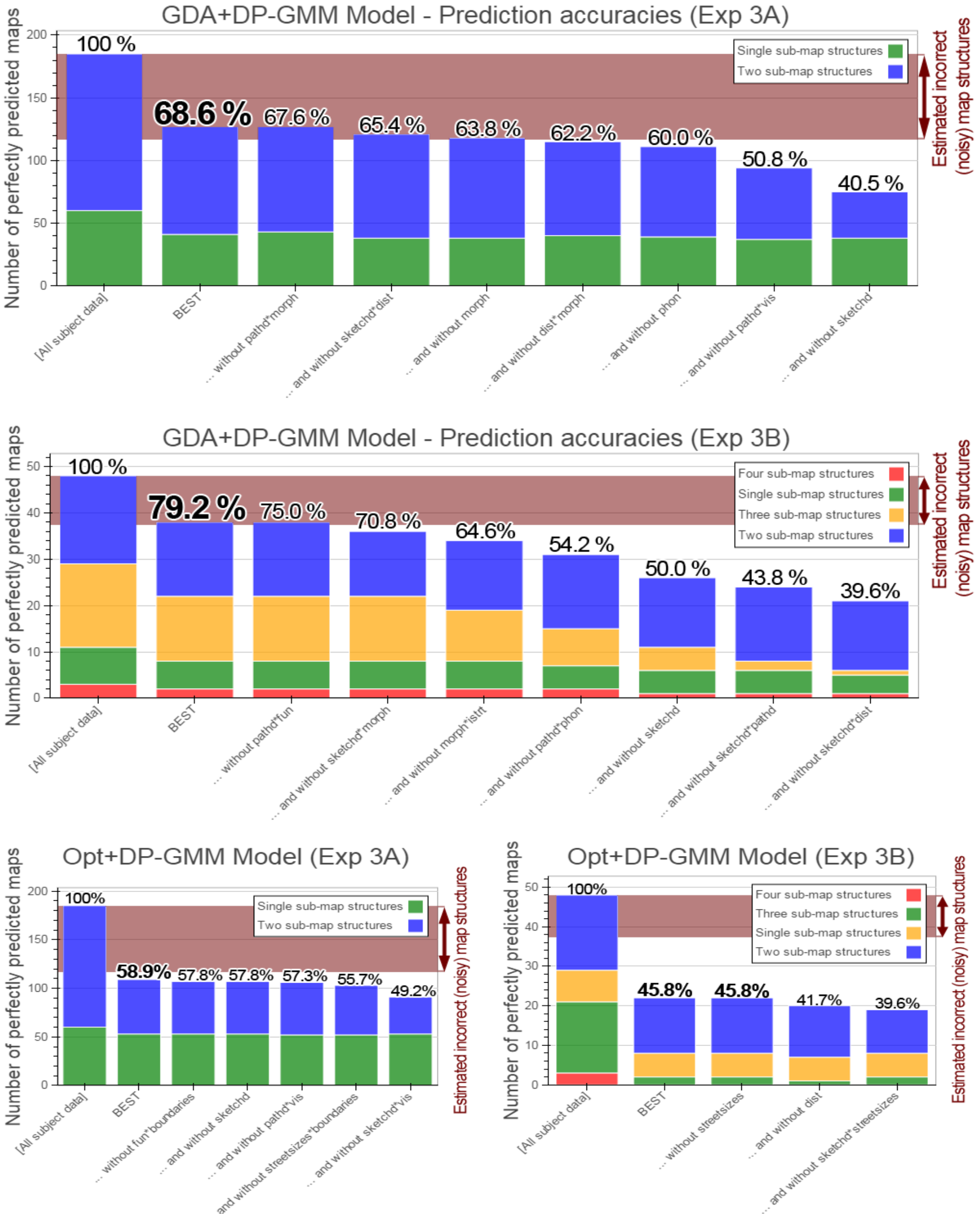


Figure 12: Accuracies obtained by predicting participant's map structures using DP-GMM clustering under the learned subject-specific models. The first bar shows the number of all subject map structures (and, within them, the numbers of structures containing the specified numbers of sub-maps). Top: results in condition A. The first bar shows all subject maps, the second the prediction accuracy for the best feature set (Euclidean, path*morphological, sketch map, path*visual, phonetic, Euclidean*morphological, morphological, sketch map*Euclidean distances); bars 3-8. show accuracies when successively removing the last feature. Middle: results in condition B. The second bar shows the best prediction accuracy (Euclidean, path*functional, sketch map*morphological, morphological*separating streets, path*phonetic, sketch map, sketch map*path and sketch map*Euclidean distances); bars 3-8. accuracies when successively removing the last feature. Bottom: Prediction accuracies of the optimization-based subject-specific model. Best model accuracies for condition A and B are shown in the second bar of the left and right bottom plot.

Even using solely features which can be objectively derived from geospatial (and linguistic) information from participants' specified buildings, without collecting any subjective data such as sketch maps or visual similarity judgements for the test maps, solid prediction is still possible - 70.8% for condition A and 68.8% in B (although recall sequences still need to be collected in order to learn subject-specific models). This makes a subject model, once learned, applicable to any environment encountered by that subject.

These results further substantiate the plausibility of the clustering hypothesis, and in particular, provide evidence that nonparametric Bayesian clustering is a suitable model not only for human category learning (Griffiths et al., 2007), but also for cognitive map structure learning; and fit in well with the growing body of evidence for 'Bayesian cognition' (Tenenbaum et al., 2011).

4. General Discussion

A growing body of evidence suggests that rather than storing spatial information within some global reference frame, human spatial memory employs local, object-centered representations (Marchette & Shelton, 2010; Chen & McNamara, 2011; Greenauer & Waller, 2010; Meilinger et al., 2014). This is consistent with the much earlier proposal that spatial memories are organized according to hierarchies (Hirtle & Jonides, 1985; McNamara, 1986; McNamara et al., 1989; Holding, 1994; Wiener & Mallot, 2003), as well as with recent neuronal evidence (Derdikman & Moser, 2010; Han & Becker, 2014).

In this paper, we made the first attempt to quantitatively explain and predict the local structure of spatial representations. We have found strong correlations between the probability that two buildings are co-represented[24] and features such as Euclidean distance, path distance, and visual and functional similarity. These correlations suggest that clustering based on proximity along these features is likely to give rise to the observed representation structure. We have developed multiple methods for exploring how important these features are for individual subjects (i.e. learning their 'psychological spaces'), even if only small amounts of data are available, and have developed and evaluated a predictive model of cognitive map structure based on Bayesian nonparametric clustering in these learned psychological spaces. We have shown that our model can successfully predict spatial representation structures in advance in the majority of cases.

The results from our model are very promising, but their plausibility depends on the empirical method used to expose spatial representation structure. Although the structures identified by our recall order paradigm are substantiated by their significant influence on several cognitive phenomena (Section 3.2), there is clearly room for improving the experimental methodology. After briefly outlining the implications of models of cognitive map structure, the discussion below outlines some alternative approaches, and suggests reasons for the imperfect prediction rates.

4.1. Implications of modelling cognitive map structure

We have reported significant effects exerted by cognitive map structure on spatial memory-related performance in Section 3.2. Together with prior evidence on priming, map distortion, distance estimation biases, and related effects, it seems clear that representation structure is relevant to spatial memory.

Apart from psychology, its investigation is also of interest for neuroscience. Strong evidence exists for hierarchies in the neural correlates of rodent spatial memory, place cells and grid cells, specialized neuron types discovered in mammalian - and, more recently, human - brains (Ekstrom et al., 2003; Jacobs et al., 2013), and is shown to play a key role in representing space (Moser et al., 2008). Place cells show increased activity in small, spatially localized areas, encoding spatial locations within particular spaces - with firing patterns changing significantly upon switching or changing immediate surroundings (the set of active place cells is completely different in separate environments). Grid cell firing shows a highly regular, triangular grid spanning the surface of an environment, independently of its configuration of landmarks, thus encoding a direction and distance metric.

Both of these spatially relevant neuron types have been observed to show natural hierarchies, with the granularity of representations (the sizes of the firing fields of individual cells) increasing from dorsal to ventral

[24] Stored on the same representation, as indicated by the recall order paradigm (i.e. always recalled together)

poles of the relevant brain areas (Brun et al., 2008; Kjelstrup et al., 2008). Furthermore, fragmentation in separate parts of an environment has also been observed in electrophysiological recordings of grid cells (Derdikman et al., 2009; Frank et al., 2000), indicating that instead of a single 'cognitive map', there a manifold of sub-maps are represented in brains (Derdikman & Moser, 2010).

However, the connection between these hierarchical and/or fragmented neural representations, and cognitive representations of map structure, remains largely unexplored. The predictive modelling approach presented in this paper could facilitate and accelerate research into this connection - after a subject-specific model has been learned from a small number of environments, subjects do not need to be subjected to arduous recall sequences (or large numbers of estimations), and can quickly be tested in large numbers of virtual reality environments in an fMRI.

Models of cognitive map structure could be of interest not only to the cognitive sciences but also to neighbouring fields. For example, in geographic information science, the insight that both planning times and estimation accuracies are improved within sub-maps compared to across, together with a subject-general model (which is good enough for this purpose - see Section 3.5), could help design schematics or transit maps which are cognitively easy to use for a majority of subjects.

Furthermore, models of human spatial representation are relevant for robotics for the purpose of communicating and interacting with humans. This is a rapidly growing area, with over three million[25] personal (non-industrial) service robots sold in 2012; a figure that can be expected to grow with the increasing demands on care robotics due to the rapid ageing of the world population. A model of spatial representation structure could allow artificial agents to use and understand human-like concepts (for example, translating latitudes and longitudes to easily understandable expressions like 'between the shopping area and the university buildings'). Approaches to conceptualize spatial representations exist only for indoor robots (Zender et al., 2008). The present approach, in contrast, is applicable to unconstrained outdoor environments (and is demonstrated by our results to work in a human-like fashion in over a hundred cities).

Finally, the particular way individual subjects structure their commonly encountered environments depending on past experience and task demands could give insight into computationally more efficient spatial representations for artificial intelligence (AI). With only around 40 million principal neurons in the human Hippocampus (Andersen et al., 2006), adults seem to be able to effortlessly store and recall navigation-relevant spatial details of many dozens of cities and hundreds of square kilometers. Storing a comparable amount on a trivial AI map representation such as an occupancy grid (Elfes, 1989), with the accuracy relevant for navigation, and including rich perceptual information, is not possible using today's hardware (let alone searching through such a vast database in split seconds, as humans are able to do). Human spatial representation structure could give inspiration for more efficient computational structures for representing space.

4.2. Alternative empirical approaches to uncovering cognitive map structure

Since humans do not have introspective insight into their own memory structure, uncovering organization principles of spatial memory is challenging. Several methods have been proposed in the literature to investigate which reference frames, or imposed structure, might be employed by participants. Of these, the recall order paradigm was used here, and described in Section 2. Its main shortcomings are the lack of robustness to outliers due to e.g. lapses of attention (mitigated by the jackknifing procedure), and the influences of phonological and morphological features of verbally cued items (mitigated by spatial cueing, as in Experiment 3). Despite these shortcomings, the structures extracted by this method have substantial influence on various cognitive phenomena, as reported in Section 3.2.

Other experimental approaches for investigating representation structure include judgements of relative direction (JRD), in which subjects imagine standing at some specified location and heading, are asked to point to specified objects they have memorized previously. The angular error in JRD seems to be strongly affected by interobject spatial relations (rather than only depending on a global reference frame), with better accuracy for judgements aligned with the intrinsic reference frame of an array of objects both in

[25] According to the World Robotics 2013 Service Robot Statistics, http://www.ifr.org/service-robots/statistics/

navigation space (McNamara et al., 2003; Meilinger et al., 2014) and in small-scale environments in a room (Mou & McNamara, 2002). These experiments have utilized object arrays with clear axes of alignment, either employing a grid-like array (mainly used in small-scale experiments) or making use of single major roads or paths as intrinsic axes in large-scale surroundings. This setup limits the applicability to general environments. However, the idea of direction judgement errors induced by changes of reference frame is generalizable, and has also been used to investigate reference frames of arrays without enforced intrinsic structure (Han & Becker, 2014; Chen & McNamara, 2011). Because direction errors are smaller within reference frames than across (Han & Becker, 2014), they could in principle be used to infer representation structure. The main disadvantage of this approach is the large number of direction estimations required to distinguish reference frames reliably, due to the large variance of direction errors. Furthermore, the number of estimations needed for pairwise comparison grows quadratically with the number of objects and / or frames (none of the cited papers compare more than two frames).

Cognitive map structure impinges on behavioural performance in several ways, most notably including biasing direction estimation (see above), distance estimation - overestimated across- and underestimated within representations (Hirtle & Jonides, 1985) -, and priming, i.e. accelerated recognition latencies (McNamara et al., 1989), direction estimation latencies (Han & Becker, 2014), and verifications of spatial relations (Hommel et al., 2000). All of these biases in errors or response times cause the same difficulties when trying to infer the exact representation structure for a particular participant - due to large variances, a very large number of judgements is required to obtain acceptable statistical significance (and the number grows quadratically with the number of objects). How to mitigate this problem, and which of these metrics have the smallest variance and thus highest reliability for map structure extraction, as well as whether they all yield consistent structures as would be expected, remain important questions for future research on cognitive map structuring.

Assuming either no distractions, or that jackknifing can successfully eliminate the majority of outliers caused by distractions, the recall order paradigm is able to provide the most deterministic way of inferring map structure, since it does not rely on comparing distributions of errors (or response times) using significance testing. It is also deterministic over time, resulting in very similar structures to the original hierarchies when re-testing subjects several weeks later (Hirtle & Jonides, 1985). These advantages, together with the difficulty of obtaining statistically significant results from error / RT patterns with high variances, have motivated our choice for the recall order paradigm for uncovering the structures modelled in this work.

4.3. Obstacles to predicting cognitive map structure

Our results indicate strong correlations of co-representation probability with distance (Section 3.3), suggesting that a clustering mechanism underlies map structures, and substantiating the plausibility of our computational model. However, these conclusions are based on a number of assumptions; and it is possible that some of them might not be correct. Below, we list some possible obstacles to a predictive model based on these assumptions.

First, it might be the case that subjects did not learn allocentric spatial representations of their chosen buildings at all. They might have painstakingly constructed the sketch maps in these experiments from egocentric representations, for example by imagining egocentric vectors from a particular vantage point, and estimating distances. If subjects can accurately estimate distances, then this procedure might yield sketch maps that are better than random, despite the absence of a metric cognitive map (subjects might well do this, for example, if they have only ever visited their chosen buildings by underground public transport). However, note that 1) in this case they would be violating the experiment instructions, which state that they need to know how to walk from any of the buildings to any of the others, and 2) it is much harder to draw accurate sketch maps when estimating from only one (or few) egocentric vantage points, as opposed to when a full 'map' is accessible allowing the choice of any building or points between buildings as vantage points.

Second, subject cognitive maps might be unstructured. However, according to the recall order paradigm, structure is evident from the recalls of a majority of subjects and subject maps. There is also the independent evidence of several distinct local reference frames, and of local neural representations (see above).

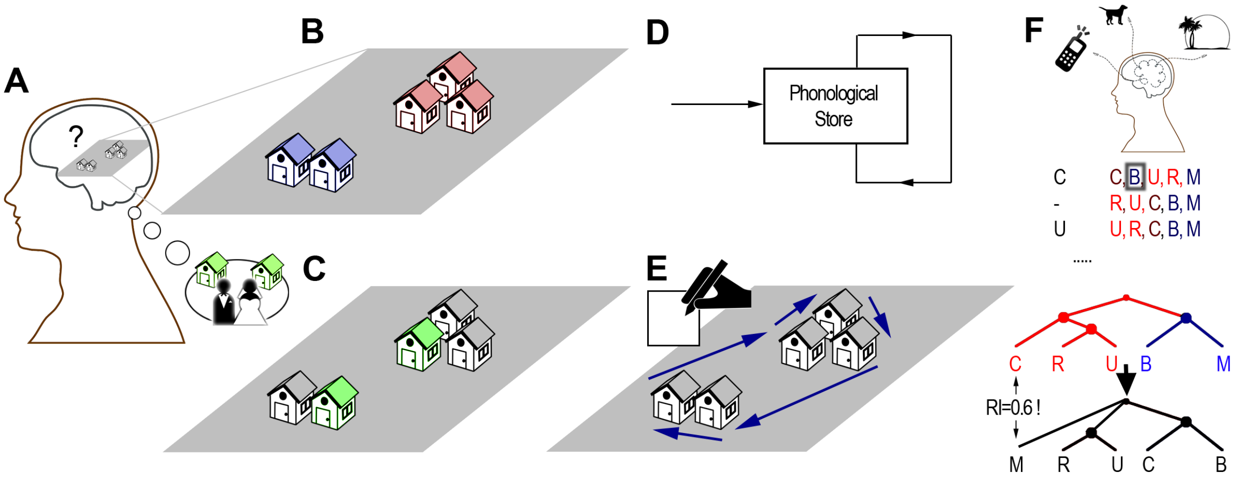


Figure 13: Possible obstacles to predicting subject cognitive map structures. A: Subjects may not have formed allocentric cognitive maps. B: Their maps may not have been structured. C: The apparent structure might be due to episodic memories, emotionally significant events, or other types of non-spatial long-term memory. D: Spurious structure might arise from articulatory rehearsal or other working memory strategies, instead of LTM. E: Subjects can list or sketch their buildings on paper, instead of recalling them from memory, to make the task faster and easier; usually resulting in circular recall sequences. F: Mind wandering or lapses in attention during recall sequences can cause tree analysis to reconstruct incorrect map structures.

Third, apparent structure might actually arise from non-spatial context effects or long-term memory events which happened at, or are relevant to, a sub-set of buildings or locations on a subject's cognitive map. For example, a subject might cluster together multiple restaurants after having had dinner at all of them with her significant other. When filling out the recall sequences, she might employ her salient episodic memories of these dinners to quickly recall these restaurants (and recall them together, which would lead to the tree analysis algorithm to assume that they are clustered together). It is difficult to exclude such influences in the real-world experiments, as most buildings familiar to subjects will have some sort of episodic memories associated with them. How frequent such influences are, and to what extent they distort apparent map structure, remain questions for future research (one approach might be trying to induce meaningful episodic memories in the virtual reality experiment, and measure their effects). However, if a majority of subject map structures had been affected by such context effects (which naturally cannot be modelled with the described features), reliable prediction would not be possible at all. The observation that a majority of structures *can* be predicted suggests that these influences affect a minority of recalled structures.

Fourth, spurious structures could appear in the recall sequences from phonetic or morphological name similarity in case subjects use articulatory rehearsal to facilitate quick recall; in which case it is a natural strategy to rehearse and recall similarly sounding object names together. This was indeed a significant influence in the verbally cued experiments (Experiment 1 and 2), although much weaker than the dominating influence of spatial distance. However, it seems that the effect can be mitigated substantially by changing the cue modality from verbal to visuospatial cues, which reduces the correlations between phonetic/morphological similarity and co-representation probability to insignificant levels (see Figure 6). A further possible objection related to working memory, that the uncovered structures might be learned during the experiment (instead of arising from long-term spatial memory), can be ruled out based on the approximately uniform distribution of outlier positions (the first few sequences were not more likely to be outliers than the last few sequences, and no evidence for any learning of map structures during the real-world experiments could be found in the data - see Supplementary Information for details).

Fifth, in the real-world experiments during which subjects were not observed, they could have lightened the cognitive load and speeded up the process by either writing down the list of buildings, or sketching a map on paper, and then reading instead of recalling. Although they were explicitly instructed to do everything from memory, without looking anything up, an unfortunate side effect of the monetary re-compensation is that they have financial incentive to speed up the task (however, (Goodman et al., 2013) have found no significant difference between the ratio of correct answers between Mechanical Turk participants and supervised subject from a middle-class urban neighbourhood; although there was a significant difference to student participants). The proportion of subjects ignoring task instructions can be reduced by ensuring that most of their other tasks were accepted by requesters on MTurk (in these experiments, they were required to have at least 95% approval rating on previous jobs to ensure higher data quality). Furthermore, since the easiest strategy when using a list or a sketch on paper is to always use the same ordering, this should cause recall sequences to be circular, which can be detected in the data. As would be expected, the rate of circular recalls is significantly higher for the MTurk subjects (Experiment 3) - 12.6% - than for the student participants of Experiment 2 - 5.3%. However, they are still a minority of the data, and have been excluded in the reported analyses (as they lead to a lack of apparent structure).

An additional obstacle to predicting cognitive map structure is the rigidity of the tree analysis algorithm. Sub-maps are only recognized as such if they occur together, without interruption, in *every single recall sequence*. Figure 13 F illustrates an example (revisiting the example from Figure 4) where a distraction, which interrupts the sequence cued with ‘C’ and causes the participant to continue with ‘B’, for example because the distraction has reminded him of ‘B’. This causes a substantially different extracted map structure - were a well-trained predictive model to predict the correct (CRU) - (BM) sub-map structure, it would show up as an incorrect prediction, and to have a Rand index of 0.6 instead of 1.0. Section 3.5 suggests a calculation of how many such such incorrectly inferred map structures there might be in our data, based on the percentage of recognized outliers using jackknifing.

Apart from devising a less simplistic outlier detection method, one possibility to reduce the occurrence of distractions - for future work - would be timing all recall sequences, and discarding those that exceed a temporal threshold, forcing participants to re-do the recall.

5. Conclusion

The way spatial memories of open, large-scale environments are structured has remained an unanswered question. In this paper, we have provided the first attempt at a quantitative answer, hypothesizing that cognitive map structure arises from clustering in some subject-specific psychological space, including (but not necessarily limited to) a list of features such as spatial distance, separating boundaries and streets, and visual and functional similarity, which we have proposed based on past empirical results. As this claim implies a strong dependence between whether or not objects are stored on the same representations, and these features, we have examined this dependence using subjects from over a hundred cities worldwide. We have found that there is a strong correlation between the probability of co-representation of buildings and their distance in these features (including, perhaps surprisingly, their visual similarities). Furthermore, we report that despite the noisy inference of subject map structures, they can be predicted correctly in a majority of cases, after learning subjects’ psychological spaces and applying clustering, using a novel computational model of cognitive map structuring based on Bayesian models of cognition. Together, these results provide strong support for the clustering hypothesis, and for the plausibility of a Bayesian model of cognitive map structuring.

Acknowledgements

This work has been supported by EPSRC (Engineering and Physical Sciences Research Council) grant EP/I028099/1, and FWF (Austrian Science Fund) grant P25380-N23.

References

Andersen, P., Morris, R., Amaral, D., Bliss, T., & O’Keefe, J. (2006). *The hippocampus book*. Oxford University Press.

Baghshah, M. S., & Shouraki, S. B. (2010). Kernel-based metric learning for semi-supervised clustering. *Neurocomputing*, *73*, 1352–1361.

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., Jeffery, K., & Burgess, N. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, *17*, 71–97.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645.

Bennett, A. T. (1996). Do animals have cognitive maps? *The journal of experimental biology*, *199*, 219–224.

Bensmail, H., & Celeux, G. (1996). Regularized gaussian discriminant analysis through eigenvalue decomposition. *Journal of the American statistical Association*, *91*, 1743–1748.

Brun, V. H., Solstad, T., Kjelstrup, K. B., Fyhn, M., Witter, M. P., Moser, E. I., & Moser, M.-B. (2008). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*, *18*, 1200–1212.

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*, 340.

Canini, K. R., Shashkov, M. M., & Griffiths, T. L. (2010). Modeling transfer learning in human categorization with the hierarchical dirichlet process. In *ICML* (pp. 151–158).

Chen, X., & McNamara, T. (2011). Object-centered reference systems and human spatial memory. *Psychonomic bulletin & review*, *18*, 985–991.

Chitta, R., Jin, R., Havens, T. C., & Jain, A. K. (2011). Approximate kernel k-means: Solution to large scale kernel clustering. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 895–903). ACM.

Cohen, G. (2000). Hierarchical models in cognition: Do they have psychological reality? *European Journal of Cognitive Psychology*, *12*, 1–36.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PloS one*, *8*, e57410.

Derdikman, D., & Moser, E. I. (2010). A manifold of spatial maps in the brain. *Trends in cognitive sciences*, *14*, 561–569.

Derdikman, D., Whitlock, J., Tsao, A., Fyhn, M., Hafting, T., Moser, M., & Moser, E. (2009). Fragmentation of grid cell maps in a multicompartment environment. *Nature neuroscience*, *12*, 1325–1332.

Ekstrom, A. D. (2015). Why vision is important to how we navigate. *Hippocampus*, .

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*, *424*, 184–187.

Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation. *Computer*, *22*, 46–57.

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*, 195.

Frank, L. M., Brown, E. N., & Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, *27*, 169–178.

Gablonsky, J. M., & Kelley, C. T. (2001). A locally-biased form of the direct algorithm. *Journal of Global Optimization*, *21*, 27–37.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.

Gershman, S. J., & Blei, D. M. (2012). A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*, 1–12.

Gibson, B. R., Rogers, T. T., & Zhu, X. (2013). Human semi-supervised learning. *Topics in cognitive science*, *5*, 132–172.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in cognitive sciences*, *5*, 236–243.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, *26*, 213–224.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*, 33–51.

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.

Griffiths, T. L., Canini, K. R., Sanborn, A. N., & Navarro, D. J. (2007). Unifying rational models of categorization via the hierarchical dirichlet process. In *Proceedings of the 29th annual conference of the cognitive science society* (pp. 323–328).

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.

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

Holding, C. S. (1994). Further evidence for the hierarchical representation of spatial information. *Journal of Environmental Psychology*, *14*, 137–147.

Hommel, B., Gehrke, J., & Knuf, L. (2000). Hierarchical coding in the perception and memory of spatial layouts. *Psychological Research*, *64*, 1–10.

Hosmer, D. W., & Lemeshow, S. (2004). *Applied logistic regression*. John Wiley & Sons.

Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., Loftus, M. M., Staskute, L., & Spiers, H. J. (2014). The hippocampus and entorhinal cortex encode the path and euclidean distances to goals during navigation. *Current Biology*, *24*, 1331–1340.

Hughes, M. C., & Sudderth, E. (2013). Memoized online variational inference for dirichlet process mixture models. In *Advances in Neural Information Processing Systems* (pp. 1133–1141).

Hurts, K. (2008). Spatial memory as a function of action-based and perception-based similarity. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1165–1169). SAGE Publications volume 52.

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X.-X., Suthana, N., Sperling, M. R., Sharan, A. D., Fried, I. et al. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature neuroscience*, *16*, 1188–1190.

Jeffery, K. J. (2015). Distorting the metric fabric of the cognitive map. *Trends in Cognitive Sciences*, .

Jeffery, K. J., & Burgess, N. (2006). A metric for the cognitive map: found at last? *Trends in cognitive sciences*, *10*, 1–3.

Jones, D. R., Perttunen, C. D., & Stuckman, B. E. (1993). Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Applications*, *79*, 157–181.

Khorsi, A. (2013). On morphological relatedness. *Natural Language Engineering*, *19*, 537–555.

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

Lewis, D. D., & Gale, W. A. (1994). A sequential algorithm for training text classifiers. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 3–12). Springer-Verlag New York, Inc.

MacKinnon, J. G. (2009). Bootstrap hypothesis testing. *Handbook of Computational Econometrics*, (pp. 183–213).

Marchette, S. A., & Shelton, A. L. (2010). Object properties and frame of reference in spatial memory representations. *Spatial Cognition & Computation*, *10*, 1–27.

Marr, D., & Poggio, T. (1976). From understanding computation to understanding neural circuitry. *AI Memo*, . URL: `http://mit.dspace.org/handle/1721.1/5782`.

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

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.

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

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*, 663–78. doi:`10.1038/nrn1932`.

Meilinger, T., Riecke, B. E., & Bülthoff, H. H. (2014). Local and global reference frames for environmental spaces. *The Quarterly Journal of Experimental Psychology*, *67*, 542–569.

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*, 69–89. doi:`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*, 162.

Nachar, N. (2008). The mann-whitney u: a test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, *4*, 13–20.

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y.-G., & Tucker, D. G. (1986). Inferring students' cognitive structures and their development using the "ordered tree technique". *Journal of Educational Psychology*, *78*, 130.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, *115*, 39.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map* volume 3. Clarendon Press Oxford.

Ong, C. S., Williamson, R. C., & Smola, A. J. (2005). Learning the kernel with hyperkernels. In *Journal of Machine Learning Research* (pp. 1043–1071).

Philips, L. (2000). The double metaphone search algorithm. *C/C++ users journal*, *18*, 38–43.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, *66*, 846–850.

Redner, R. A., & Walker, H. F. (1984). Mixture densities, maximum likelihood and the em algorithm. *SIAM review*, *26*, 195–239.

Reineking, T., Kohlhagen, C., & Zetzsche, C. (2008). Efficient wayfinding in hierarchically regionalized spatial environments. In *Spatial Cognition VI. Learning, Reasoning, and Talking about Space* (pp. 56–70). Springer.

Reitman, J. S., & Rueter, H. H. (1980). Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology*, *12*, 554–581.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In *Proceedings of the 28th annual conference of the cognitive science society* (pp. 726–731).

Sato, N., & Yamaguchi, Y. (2009). Spatial-area selective retrieval of multiple object–place associations in a hierarchical cognitive map formed by theta phase coding. *Cognitive neurodynamics*, *3*, 131–140.

Settles, B. (2010). Active learning literature survey. *Computer Sciences Technical Report 1648*, .

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

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.

Spelke, E., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive Science*, *34*, 863–884.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, *331*, 1279–1285.

Thomas, R., & Donikian, S. (2007). A spatial cognitive map and a human-like memory model dedicated to pedestrian navigation in virtual urban environments. In *Spatial Cognition V Reasoning, Action, Interaction* (pp. 421–438). Springer.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*, 189.

Voicu, H. (2003). Hierarchical cognitive maps. *Neural Networks*, *16*, 569–576.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in cognitive sciences*, *6*, 376–382.

Wiener, J. M., & Mallot, H. A. (2003). 'fine-to-coarse'route planning and navigation in regionalized environments. *Spatial* *cognition and computation*, *3*, 331–358.

Xing, E. P., Jordan, M. I., Russell, S., & Ng, A. Y. (2002). Distance metric learning with application to clustering with side-information. In *Advances in neural information processing systems* (pp. 505–512).

Yang, L., & Jin, R. (2006). Distance metric learning: A comprehensive survey. *Michigan State Universiy*, *2*.

Zender, H., Mozos, O. M., Jensfelt, P., Kruijff, G.-J., & Burgard, W. (2008). Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, *56*, 493–502.