1

1

2

Finding landmarks – An investigation of viewing behavior during spatial navigation in VR using a graph-theoretical analysis approach

6

by Jasmin L. Walter[1,¶,*], Lucas Essmann[1,¶], Sabine U. König[1], Peter König[1,2]

8
9   1 Institute of Cognitive Science, University of Osnabrück, Osnabrück, NI, Germany
10 2 Department of Neurophysiology and Pathophysiology, University Medical Center
11 Hamburg-Eppendorf, Hamburg, HH, Germany
12
13
14 * Corresponding author:
15 jawalter@uni-osnabrueck.de (JLW)
16

17 ¶ These authors contributed equally to this work.

18

19 **Key words: Spatial Navigation, VR, Graph Theory, Eye Tracking, Landmarks**

20

21

22

Abstract

Vision provides the most important sensory information for spatial navigation. Recent technical advances allow new options to conduct more naturalistic experiments in virtual reality (VR) while additionally gather data of the viewing behavior with eye tracking investigations. Here, we propose a method that allows to quantify characteristics of visual behavior by using graph-theoretical measures to abstract eye tracking data recorded in a 3D virtual urban environment.

The analysis is based on eye tracking data of 20 participants, who freely explored the virtual city Seahaven for 90 minutes with an immersive VR headset with an inbuild eye tracker. To extract what participants looked at, we defined "gaze" events, from which we created gaze graphs. On these, we applied graph-theoretical measures to reveal the underlying structure of visual attention.

Applying graph partitioning, we found that our virtual environment could be treated as one coherent city. To investigate the importance of houses in the city, we applied the node degree centrality measure. Our results revealed that 10 houses had a node degree that exceeded consistently two-sigma distance from the mean node degree of all other houses. The importance of these houses was supported by the hierarchy index, which showed a clear hierarchical structure of the gaze graphs. As these high node degree houses fulfilled several characteristics of landmarks, we named them "gaze-graph-defined landmarks". Applying the rich club coefficient, we found that these gaze-graph-defined landmarks were preferentially connected to each other and that participants spend the majority of their experiment time in areas where at least two of those houses were visible.

Our findings do not only provide new experimental evidence for the development of spatial knowledge, but also establish a new methodology to identify and assess the function of landmarks in spatial navigation based on eye tracking data.

Author Summary

The ability to navigate and orient ourselves in an unknown environment is important in everyday life. To better understand how we are able to learn about a new environment, it is important to study our behavior during the process of spatial navigation. New technical advances allow us to conduct studies in naturalistic virtual environments with participants wearing immersive VR-headsets. In addition, we can use eye trackers to observe the participant's eye movements. This is interesting, because observing eye movements allows us to observe visual attention and therefore important cognitive processes. But, it can be difficult to analyze eye tracking data that was measured in a VR environment, as there is no established algorithm yet. Therefore, we propose a new method to analyze such eye tracking data. In addition, our method allows us to transform the eye tracking data into graphs which we can use to find new patterns in behavior that were not accessible before. Using this methodology, we found that participants who spend 90 min exploring a new virtual town used some houses as orientation anchors which we call gaze-graph-defined landmarks. Our further analysis revealed also new characteristics of those houses that were not yet associated with landmarks.

Introduction

64

65 Having a sense of orientation and being able to navigate in the world that surrounds us is
66 essential in everyday life. Specifically, the awareness of the own position in space combined
67 with the ability to remember key locations to plan mental routes between them (1) is crucial.
68 This enables efficient navigation to a location by using globally accessible knowledge of a new
69 environment or previously acquired knowledge of a known environment. Overall, the ability
70 to remember and use important locations and their relations is essential for spatial navigation.

71   In classical research, spatial navigation depends on three types of knowledge (2). First,
72 landmarks are characterized by their salience against their surrounding (3). Consequently,
73 they may serve as anchors for localization and are memorized when exploring a new
74 environment, thus forming landmark knowledge (2,4,5). Second, route knowledge refers to
75 the ability to travel along paths and remember routes between landmarks. It is thought to be
76 acquired during active navigation (6), but does not necessarily contain metric distance or
77 direction information (4). Third, survey knowledge is described as a mental map-like
78 representation developed by assembling landmark and route knowledge (4) including spatial
79 relations between multiple landmarks (7,8). Additionally, survey knowledge is thought to
80 contain information about metric properties and the relation to cardinal direction (9). Overall,
81 knowledge of landmarks, routes and cognitive maps have been cornerstones for spatial
82 navigation research for decades.

83   In spite of their ubiquitous use, the concept of landmarks, routes and cognitive maps
84 does not come without problems. Often, landmark based navigational learning is investigated
85 in environments where single cues are introduced to serve as landmarks in an otherwise
86 undifferentiated environment. For example, the Morris water-maze is a widely used task to
87 study the physiological mechanisms of spatial learning in rodents (10–12). Similarly,
88 adaptations of the Morris water-maze in virtual environments (13–15) and other maze tasks
89 (16,17) have been used to investigate spatial navigation in humans. Moreover, performing eye
90 tracking allows investigation of visual interaction with landmarks (13,16,18,19). Other studies
91 investigating spatial navigation were conducted in real world natural environments using
92 mobile eye tracking systems (20–25). However, with recent technological advances, the virtual
93 environments used in spatial navigation research have become more complex and naturalistic
94 (18,19). Some studies push for even more naturalistic virtual environments in combination
95 with head mounted virtual reality headsets (6,26,27). This development creates the need to
96 clearly define the concepts of landmarks, route knowledge and survey knowledge in such
97 naturalistic environments.

98   The increasing use of more realistic VR environments with freedom to move, creates
99 new challenges for the analysis of eye tracking data. In classical visual exploration of static 2D
100 images, fixations and saccades dominate. In a complex 3D environment vestibular-ocular
101 reflexes and pursuit movements additionally occur on a regular basis. However, there is no
102 established algorithm to differentiate this expanded set of eye movements in eye tracking
103 data collected in a 3D environment. Thus, to identify characteristics of visual behaviour during
104 free exploration of a virtual village and how they relate to spatial navigation, a new method is
105 needed.

4

106   In this study, we propose a procedure to analyze eye tracking data and a data driven
107 method to objectively define and quantify visual behavior with respect to spatial information.
108 We use a graph-theoretical approach to access global navigation characteristics across
109 participants and investigate the occurrence, connectivity, and navigation function of a subset
110 of houses, consistently outstanding in their graph-theoretical properties. The approach is
111 applied to eye tracking data recorded during exploration of a virtual town. Overall, our
112 findings establish a new methodology to identify and assess the function of outstanding
113 houses in spatial navigation based on eye tracking data.

Results

115 We collected eye tracking data of 22 participants (11 females, age: M = 22,86, SD = 2,59)
116 during 90 min of free exploration in the virtual city Seahaven. All participants gave their
117 written informed consent to participate. The study was approved by the Ethics Committee of
118 the Osnabrück University following the Institutional and National Research Committees'
119 ethical standards. The participants wore a VR headset with an inbuild eye tracker. They moved
120 using a controller and physically rotated their body on a swivel chair to turn in the virtual world
121 (Fig. 1a). The virtual city was built on an island and comprised 213 houses (Fig. 1b). Colliders,
122 i.e., transparent closely fit box-like structures, surrounded all houses, trees, and roads. We
123 calculated the viewing direction based on the position of the participant in the VR, the rotation
124 of the headset, and information from the eye tracker. After data collection, we cast a virtual
125 ray in the viewing direction until it hit a collider, indicating that the participant viewed the
126 respective object (28). This process was completed 30 times per second. Aggregated over all
127 participants, this led to about 3.500.000 hit points that built the basis for our data analysis.
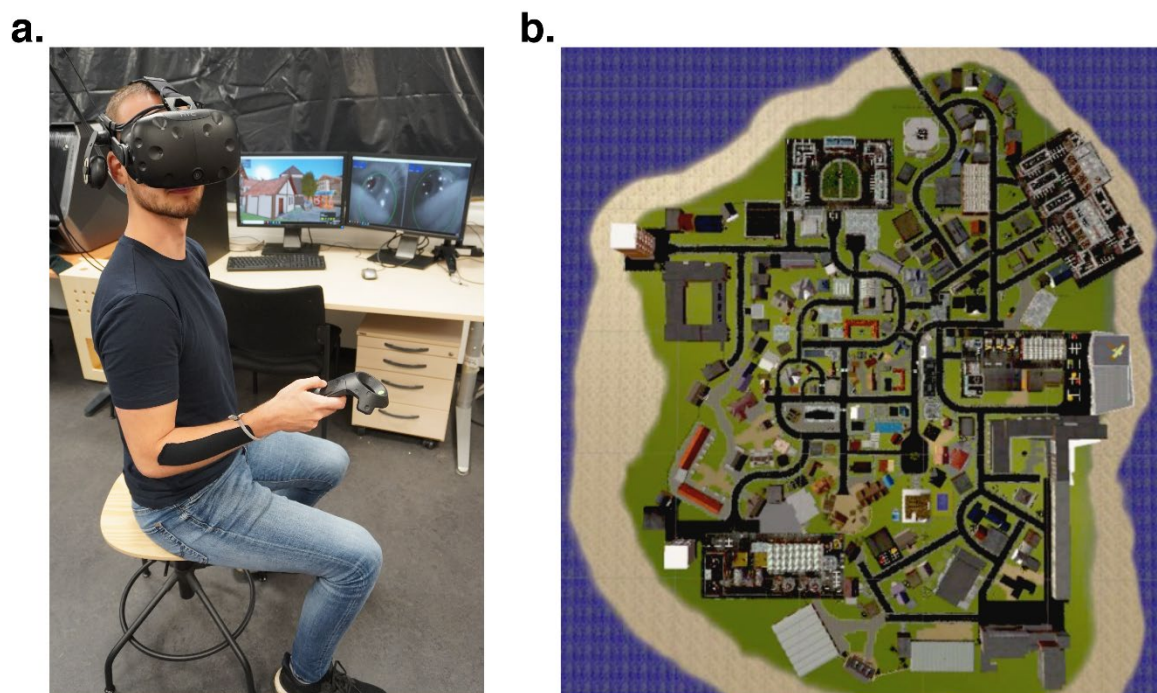
a.

b.

128 ***Figure 1: Experimental Setup and Seahaven (a)****The subject sits on a fully rotatable swivel*
129 *chair wearing headset and controller. The experimenter's screens in the background allow*

130 *monitoring the subject's visual field of Seahaven (left) and the pupil labs camera images (right).*
131 *(b) The island of Seahaven in aerial view.*

Defining gazes in VR

133 Eye tracking in an VR environment provides several challenges for data quality. Compared to
134 conventional desktop settings, VR environments allow increased freedom of movement.
135 Under these conditions, subjects perform fixations, saccades, vestibulo-occular reflexes, and
136 smooth pursuit movements. No general algorithm to classify these types of eye movements
137 has been established yet. Thus, the development of appropriate processing algorithms for the
138 eye tracking data is crucial.

139 As a first step, we use the collider hit points to identify the object the participants
140 focused their gaze on. Each collider corresponds to an entire object; thus, each hit point
141 identifies an entire object. For our analysis, houses form the regions of interests (ROIs). The
142 no house category (NH) summarizes all other collider hits except houses, e.g. grass, roads,
143 trees, and the water. Samples that do not hit a collider, thus hitting the sky or the sun, are
144 identified as "sky" category. We then combine directly consecutive hit points on the same
145 collider to identify clusters. Please note that the clusters do not contain information on where
146 in the participant's visual field the viewed object occurred. This pre-processed data of
147 combined consecutive hit points on the same collider serves as the basis for all following data
148 processing.

149 As a second step, we address the problem of missing data in individual subjects. We
150 label all data samples that were recorded with a probability of correct pupil detection of less
151 than 50% as "missing data samples". Subsequently, we exclude two participants who had
152 more than 30% of their eye tracking data classified as missing data. All further data analysis is
153 conducted with the data of the remaining 20 participants.

154 In the following we will need to differentiate between periods where participants
155 could perceive the visual stimuli or not. Given that vestibulo-occular reflexes and smooth
156 pursuit movements stabilize the retinal image in dynamic situations and allow perception in
157 that period similar to fixations (29), here and in all following analysis we subsume these under
158 the general term of fixations. In contrast, participants are blind to visual input during a
159 saccadic suppression (29,30). While classical fixation detection algorithms often differentiate
160 eye movements based on velocity, these eye movements also display a temporal disparity.
161 Specifically, saccades usually range from a duration of 10 to 100 ms, while fixations typically
162 occur from 150 to 600 ms with a mode around 250 ms (29). Therefore, we conjecture that
163 with an appropriate temporal threshold, specifically the time scale of 266 ms which is
164 equivalent to eight data samples, it is possible to identify data clusters containing fixations.

165 However, since the data still includes a considerable amount of missing data points,
166 we expect a significant number of clusters to be "cut" by missing data points, thus appearing
167 to be of shorter duration. Consequently, using a fixed temporal threshold to identify clusters
168 containing fixations creates the problem of falsely failing to identify these "cut" clusters as
169 clusters containing fixations. Therefore, to counteract this effect, it is crucial to interpolate
170 missing data samples if possible.

6

171   Therefore, as a third step in our pre-processing algorithm, we interpolate short
172   intervals of missing data. If seven or fewer data points are missing, it is improbable that
173   subjects had enough time to make a saccade to a different spatial area (ROI), finish a fixation,
174   and make a saccade back to the same area. Consequently, we interpolate data if seven or
175   fewer consecutive data points are missing and only if these occur between two clusters on the
176   same collider (ROI) (Fig. 2a). In case these interpolation criteria apply, the interpolated data
177   points are labelled with the collider name and combined with the two clusters surrounding
178   the short interval of missing data. Larger gaps are not interpolated but treated as missing data.
179   Furthermore, missing data points occurring between clusters of different colliders are not
180   interpolated, independently of the duration of the gap. This procedure ensures to capture
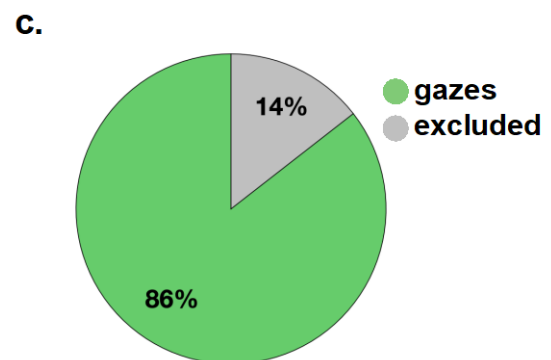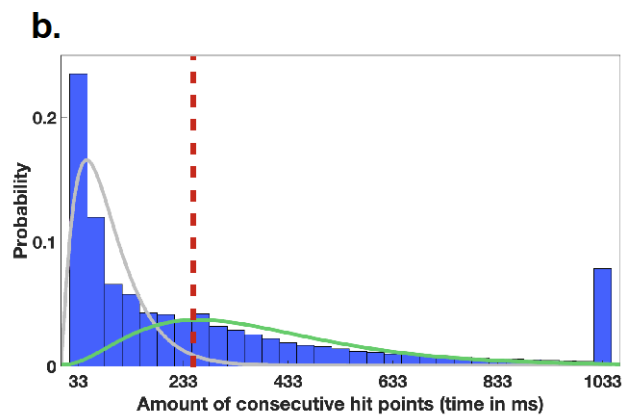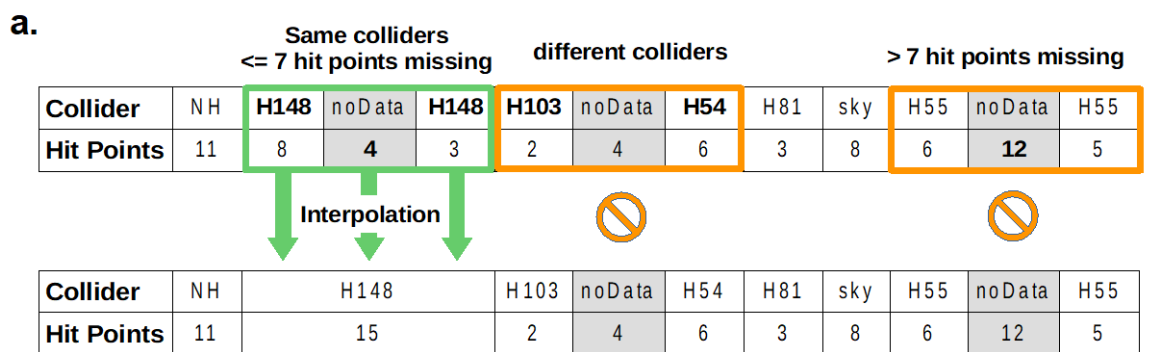181   most fixations while minimizing false interpretations of missing data.

a.

		Same colliders <= 7 hit points missing			different colliders					> 7 hit points missing		
Collider	NH	H148	noData	H148	H103	noData	H54	H81	sky	H55	noData	H55
Hit Points	11	8	4	3	2	4	6	3	8	6	12	5

Interpolation

Collider	NH	H148			H103	noData	H54	H81	sky	H55	noData	H55
Hit Points	11	15			2	4	6	3	8	6	12	5

b.

c.

14% — gazes / excluded
86%

182
183   ***Figure 2: Defining gazes. (a)** We interpolate of missing data only if no more than 7 samples*
184   *are missing consecutively (pupil detection with less than 50% probability) and if these samples*
185   *occur between two clusters on the same collider. During the interpolation process, these*
186   *samples are then unified with the two surrounding clusters to form a new cluster on the same*
187   *collider. The first row shows three clusters of missing samples (marked noData), while the*
188   *second row represents the result of the algorithm. In the first cluster (green box), both*
189   *interpolation conditions apply: there are no more than seven consecutive missing data samples*
190   *(#4) and they are surrounded by two clusters on the same collider. Consequently, these*
191   *missing samples are interpolated and combined to a new cluster. In the second cluster (1st*
192   *orange box) the first interpolation condition applies (#noData samples <= 7) but the cluster*
193   *occurs between clusters on two different colliders (H103 and H54). Therefore, no interpolation*
194   *is performed. In the third cluster (2nd orange box) only the second interpolation condition*
195   *applies. Even though the missing data samples occur between two clusters on the same*

196 *collider (H55), the first interpolation condition is violated (#noData samples > 7). Consequently,*
197 *no interpolation is performed. **(b)** Histogram of hit point cluster length distribution after*
198 *interpolation. For better visualization all cluster durations longer than 1000 ms are combined*
199 *in the last bin. The ordinate corresponds to the probability of each duration. Since previous*
200 *work used gamma distributions to model the distribution of fixation durations or response*
201 *latencies (31,32), we model the two partly overlapping gamma distributions for visualization*
202 *only, fitting the distributions of the duration of fixations (green) and non-fixation events (grey).*
203 *The dashed red line marks the separation threshold for gazes. **(c)** The pie chart shows the*
204 *result of the gaze classification across all participants.*
205

206　　　　As a fourth step, we finally apply the temporal threshold to identify clusters that
207 contain fixations during which the participants could process the visual input. As described
208 above, we adopt a threshold of eight samples, i.e. 266 ms, to identify clusters that contain at
209 least one fixation (Fig. 2b). Using this approach, shorter clusters likely to be caused by samples
210 during saccades, i.e., periods during which perception was suppressed, would be excluded
211 from further analysis. In the following, we will define these clusters containing fixations as
212 new meaningful "gaze" events and they will form the basis of our further analysis.
213　　　　This procedure classifies on average 86% of the data samples as belonging to gazes
214 (Fig. 2c). Previous studies reported that humans spent approximately 90% of viewing time on
215 fixations (29). This indicates that our attempt to capture gaze events under dynamic
216 conditions in VR is on the conservative side.

How to create graphs from gazes

218 To capture information gathered by the participants during exploration of the virtual town,
219 we create gaze graphs based on the gaze events. In these graphs each node represents a
220 house. Viewing two houses in the VR environment in direct succession gives information on
221 their relative spatial location. Therefore, if anytime during the experiment a gaze event on a
222 house is directly followed by a gaze event on another house the respective nodes are
223 connected by an undirected edge. All gaze graphs are undirected and unweighted. Hence,
224 they do not contain information about the directionality or frequency of their edges (i.e.
225 direction and frequency of succession of gaze events). The completed gaze graphs capture the
226 spatial information obtained by the visual exploration in the virtual town.
227　　　　The first 30 seconds of one participant exploring Seahaven serve as an example of the
228 process to create a graph (Fig. 3a). The graph contains one node for each house (ROI), but
229 does not consider the spatially unspecific no-house and sky category. The graph creation
230 process starts with an edge between the first house viewed and the second house viewed (Fig.
231 3b). If the gaze on a house is followed by a missing data cluster, then no edge is created. As
232 stated above, the edges are unweighted, i.e. binary. That is, if the participant looked back to
233 two houses already seen in sequence, an edge between these houses is already in place and
234 the graph is not changed. The process of edge creation is iterated for all gaze transitions.
235 Whenever, a new house is viewed, a new node is created in the graph. Fig. 3c shows a
236 visualization of the result of applying this procedure to the complete data of the example
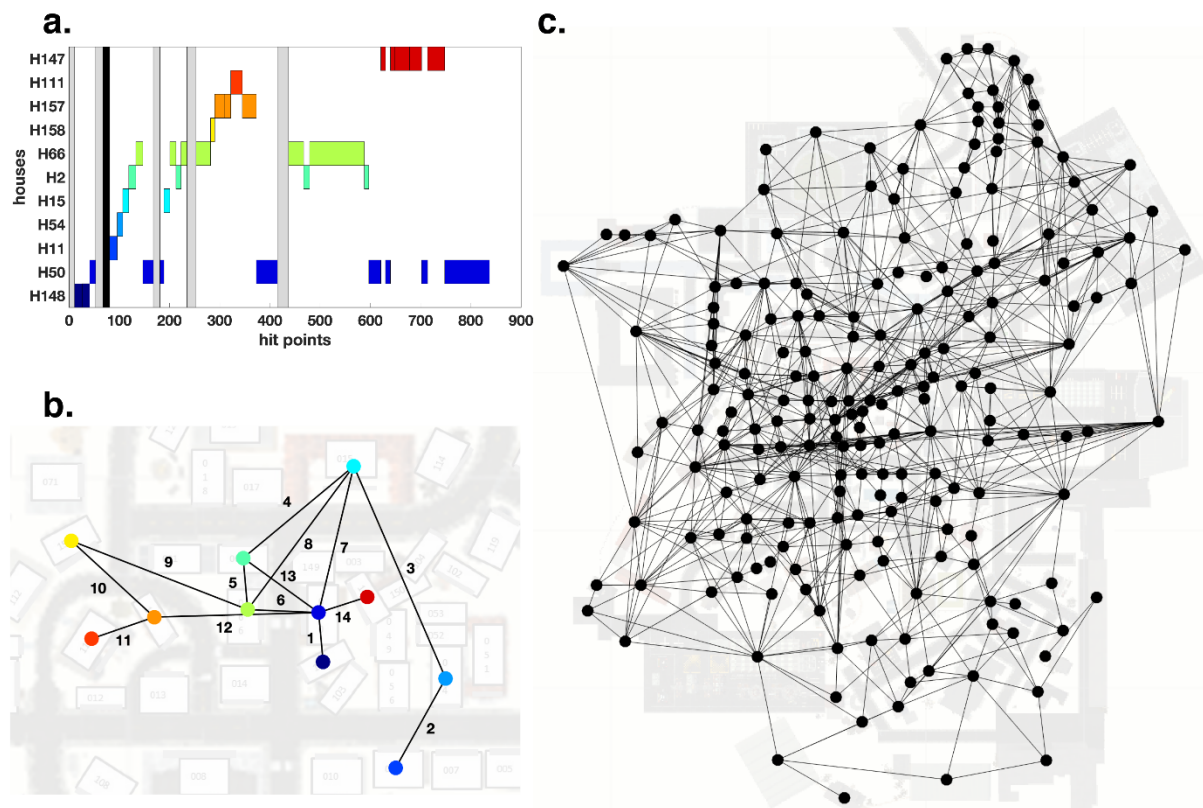237 participant.

239 **Figure 3: Graph creation. (a)** *Time line of gaze events by a participant. The abscissa*
240 *represents the first 30 seconds (900 hit points) of the recordings. The ordinate contains all*
241 *viewed houses viewed during that time line. We number houses and name them accordingly,*
242 *e.g., H148 for house number 148. In this panel each house has a distinct color for visualization*
243 *only. The grey bars represent clusters of the NH category, which are not considered during*
244 *graph creation. The black bar identifies a remaining cluster of missing data samples.*
245 *Therefore, no edge will be created at this moment in the time line. (b) The graph corresponding*
246 *to the time line of panel A is visualized on top of the map of Seahaven. The colors of the nodes*
247 *match the colors of the boxes in panel A. Edges are labelled according to the order they were*
248 *created. (c) The complete graph of a single participant based on all gaze events during 90 min*
249 *of exploration visualized on top of the 2D map of Seahaven. Note that in this visualization the*
250 *locations of the nodes correspond to the locations of the respective houses they represent in*
251 *Seahaven, however, this locational information is not contained in the graph itself.*

Is it a single city or multiple suburbs? – Graph partitioning

253 To address questions on spatial cognition, we are interested whether the Seahaven should be
254 treated as a loosely connected set of suburbs or as a coherent single city. The search for
255 distinct clusters in the graph directly relates to the problem of graph partitioning. In the field
256 of graph theory, partitioning is a well investigated problem as it divides a graph into smaller
257 mutually non-overlapping subgraphs. Of specific interest are approaches that maximize
258 within-cluster connections (those that are maintained) and minimize between-cluster
259 connections (those that are cut). Thus, we use graph partitioning to classify Seahaven as a set
260 of suburbs or one coherent city.

9

261   For graph partitioning, we employ the spectral graph analysis. This approach includes
262   three steps: the calculation of the graph's Laplacian matrix, finding the second smallest
263   eigenvalue, and splitting the graph based on the corresponding eigenvector. First, we consider
264   the adjacency matrix and the degree matrix of a graph. The former is a binary square matrix
265   where each entry indicates whether an edge connects the two nodes / houses. The latter a
266   square matrix that contains the degree of each node on its diagonal and is zero otherwise. We
267   calculate the Laplacian matrix by subtracting the degree- from the adjacency matrix. Second,
268   according to (33,34), the spectrum of a graph, i.e., the eigenvalues of the Laplacian matrix can
269   be used as a measure of graph connectivity. We calculate these eigenvalues μ_i and sort them
270   in increasing size. They have the properties that $\mu_1 = 0$ and $\mu_2 \geq 0$. The second property is valid
271   only if the graph is connected. As we eliminate all nodes/houses that were never viewed, this
272   is given in our application ($\mu_2 = 0.300+/-0.097$). Third, we consider the eigenvector with the
273   second smallest eigenvalue and map each node onto the corresponding entry. Sorting the
274   adjacency matrix in this way, a modular structure of the graph would be visible in form of a
275   block structure (Fig. 4a), which, however, is not the case in the present analysis. The
276   partitioning of the graph is performed by splitting of the eigenvector into positive and negative
277   parts (Fig. 4b) and assigning the corresponding nodes to the two separate groups. The larger
278   the gap between positive and negative values, the fewer inter-cluster connections are
279   present. In the present analysis, hardly any gap is visible. Fig. 4c visualizes the results of the
280   graph partitioning by spectral analysis with the nodes of each cluster color coded and plotted
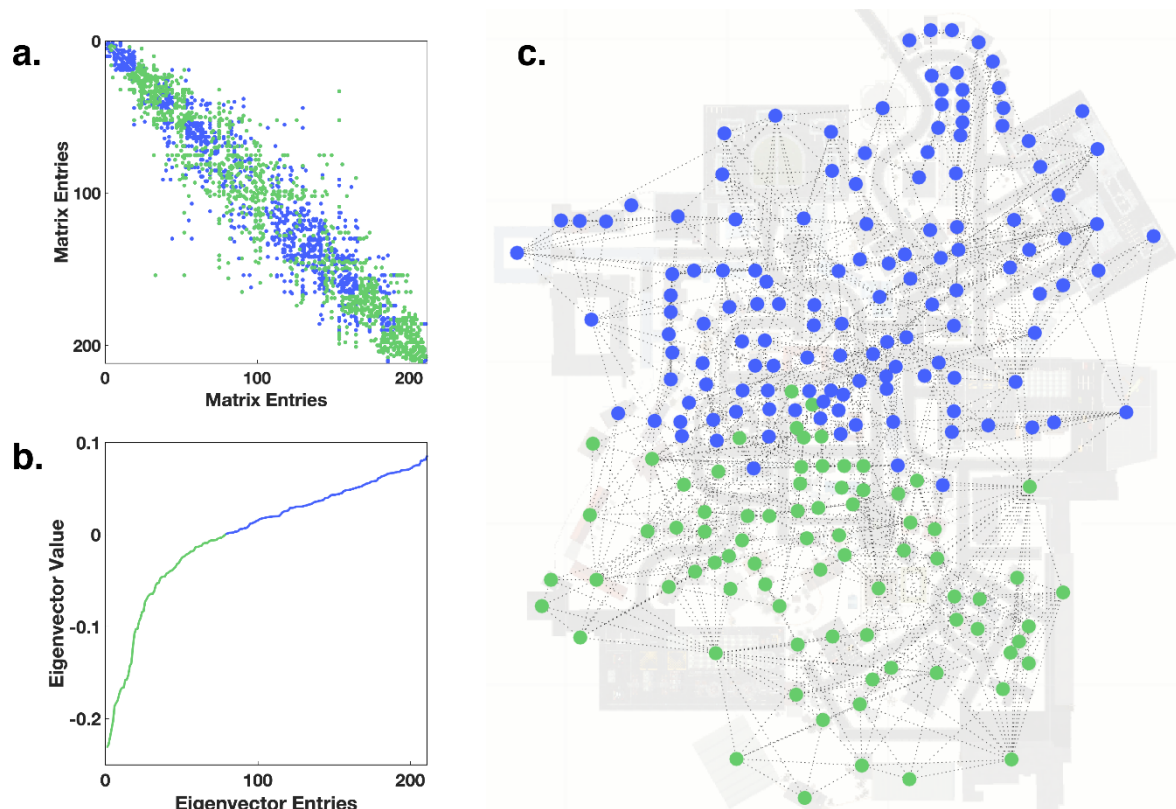281   onto the city map.

282   *Figure 4: Graph partitioning. (a) The sparsity pattern of the graph's adjacency matrix sorted*
283   *by its second smallest eigenvector. Color coded into two clusters obtained by the positive and*
284   *negative parts of the eigenvector. (b) The second smallest eigenvector of the adjacency matrix*

10

285   *is sorted ascendingly and color coded into two clusters.* **(c)** *The two clusters are displayed onto*
286   *the map of subject 35.*

287

288   To investigate the quality of the partitioning, we follow the definitions used in
289   Schaeffer (2007). As a measure of cluster goodness, we consider the intra-cluster density and
290   the inter-cluster density. The density of a graph is defined as the ratio of instantiated edges
291   relative to the number of possible edges:

292
$$\delta(G) = \frac{m}{\binom{n}{2}}$$

293   with $n = |V|$ as the number of nodes, and $m = |E|$ as the edge count. Furthermore, we
294   define the intra-cluster density as the average of the densities of both clusters and the inter-
295   cluster density as the ratio of inter-cluster edges to the maximum possible inter-cluster edges.
296   We analyze the graphs of individual participants based on 90 min exploration time. On
297   average, the participants' graphs contain 883 edges. The mean density of the graphs is
298   0.041+/-0.006. This means that on average 4.1% of all possible edges are instantiated.
299   Furthermore, dividing the graphs into two parts would require on average a cut of at least
300   9.3% of the edges. Specifically, a sufficient cut has to have 82.6 edges resulting in two clusters
301   with approximately 400 edges each. After partitioning, the mean intra-cluster density was
302   0.079+/-0.012, while the mean inter-cluster density was 0.0083+/-0.0020. These numbers
303   indicate that the graph cannot be easily partitioned, without cutting a fair number of edges.
304   In conclusion, our results reveal that the graphs cannot be distinguished into large-
305   scale clusters. That is, the exploration of Seahaven does not show separate city blocks, but
306   was rather well-balanced. Thus, the virtual environment can be treated as one coherent city.

307 **The distribution of gazes on houses - Node degree centrality**

308 We characterize the role of different houses during visual exploration by indices adapted from
309 graph theory. The node degree centrality is the main and most basic graph-theoretical
310 measure in graph-theoretical research. It is defined as the sum of all edges connecting a node:

311
$$c(i) = \sum_{j}^{N} x_{ij}$$

312 with i being the node under consideration and x_{ij} the adjacency matrix. Here, the node degree
313 of a house reflects the number of other houses a participant made a gaze to or from the house
314 under consideration in direct succession. The node degree centrality in the gaze graph thus
315 differentiates houses according to their importance during visual exploration in the virtual
316 town.

317   First, we use the node degree centrality to investigate the viewing behavior on an
318 individual participant level. Fig. 5a shows the gaze graph of a participant with individual nodes
319 color coded according to the respective node degree. Whereas many nodes have a degree
320 centrality in the single digit range, a few houses reach rather high values. The variance of node
321 degree for each subject ranged between 0 and 33. It is apparent that the range of node
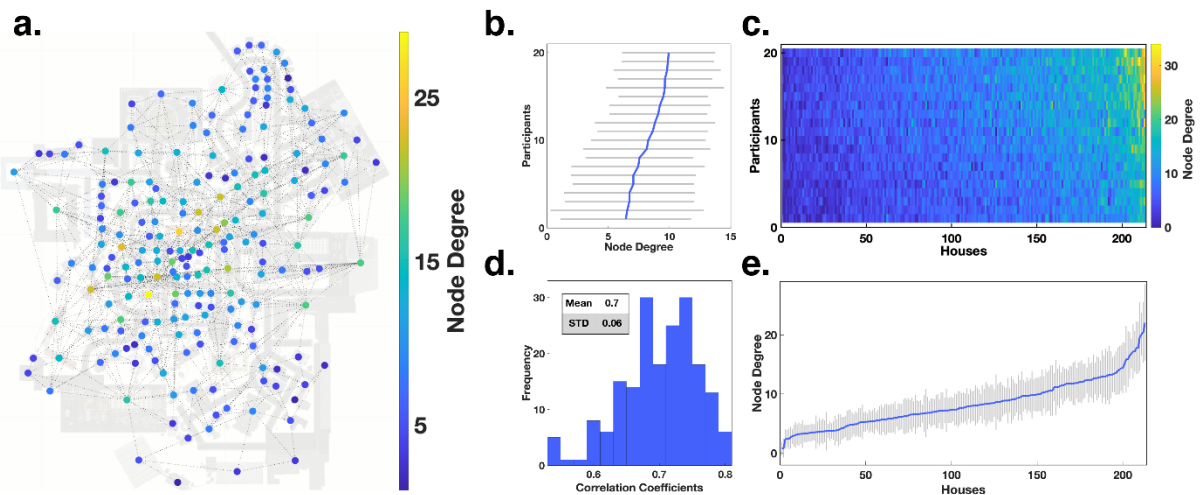322 degrees is surprisingly large.

11

Figure 5: Node degree centrality. (a) *The graph of one participant is visualized on top of the map of Seahaven. The nodes were colored according to their respective node degree centrality.* **(b)** *The mean node degree of all subjects (blue line) and their respective standard deviation (grey lines), sorted such that the average node degree increases along the ordinate.* **(c)** *A pseudo 3D plot color coding the node degree of every house (abscissa) for every subject (ordinate). The houses are sorted so that the average node degree value increases along the abscissa. Similarly, the participants are sorted, so that the average node degree increases along the ordinate. The marginals of this plot result in the panels b and e.* **(d)** *The distribution of the pairwise inter-subject correlation coefficients of the node degree values of all houses.* **(e)** *The mean node degree of each house sorted according to the mean node degree along the abscissa (blue line) and their respective standard deviations (grey lines).*

Next, we investigate the similarity of the node degree centrality distribution over subjects. We calculate the average node degree of each participant across all houses (Fig. 5b), which reveals limited variations only. The average node degree of each house across all participants (Fig. 5e) showed a monotonous linear increase. Both aspects are also combined in an image-scale visualization (Fig. 5c) that matches the order of the houses and the order of participants. Sorting the houses according to the average node degree centrality leads to a near monotonous increase on the individual participant level as well. This was quantified by the high correlation of 0.70 (+/- 0.06) of node degree centrality between participants (Fig. 5d). Thus, the node degree centrality in the gaze graph varies considerably while the values of individual houses are rather consistent across participants.

The distribution of node degree centrality over houses reveals interesting aspects. Over a large range of houses the average node degree centrality increases only slowly. Only for the last few houses, we observe a steep increase. These houses stand out from the other houses and are viewed directly before or directly after viewing many other houses. Therefore, the high node degree centrality houses may serve as important reference points.

In summary, the node degree is a simple yet powerful centrality measure that can be used to identify important nodes in the gaze graph, hence important houses in visual behavior. First results indicate that a small number of houses show an especially high node degree across all participants, setting themselves apart from the rest of the city. Interestingly, a small

354 number of houses with subject independent high node degree values, i.e. houses with many
355 visual connections to other houses, would also be the characteristics that we would expect
356 landmarks to display in gaze graphs.

High node degree centrality houses – Hierarchy index

358 o further investigate the houses with high node degree in respect to their distribution, we
359 apply the graph-theoretical measurement of the hierarchy index. This index characterizes
360 hierarchical configurations within networks. By applying it to the degree values above the
361 subject's respective median, we focus the index on the upper tail of the distribution of node
362 degree centrality. Here, we fit in a bi-log plot of the node degree centrality with frequency
363 against the node degree values (Fig. 6a), the slope of the regression line starting from the
364 subject's respective median. For the example participant reported already above this results
365 in a slope of -2.6349. Performing this analysis for all participants results in a slope smaller than
366 -2.00 throughout, with a mean of -2.91. Furthermore, the small standard deviation of 0.34
367 showed, that the hierarchy index is similar across subjects (Fig. 6b). According to (35),
368 networks with strong hierarchical configurations, i.e., with many low degree nodes and few
369 high degree nodes, had a slope of below -2 (or above 2). Therefore, the hierarchy index reveals
370 a clear hierarchical structure of the gaze graphs and emphasizes the importance of the few
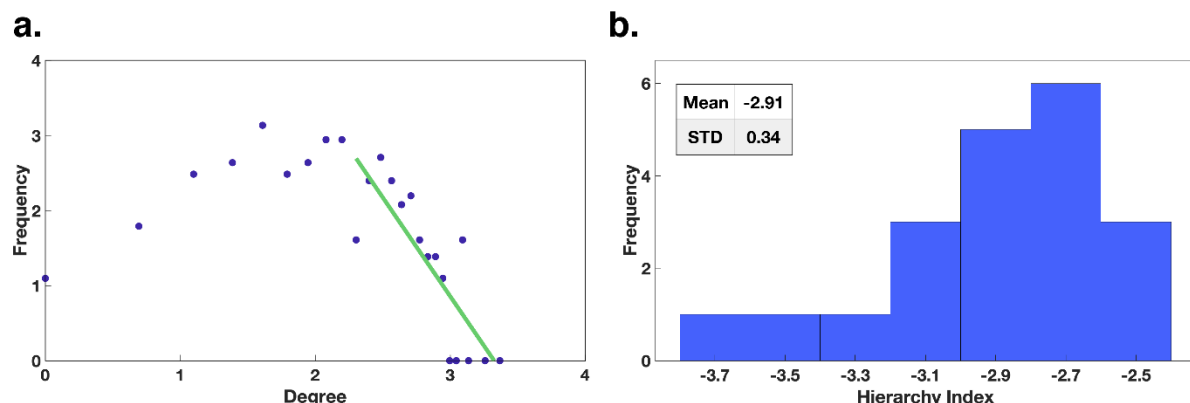371 exceptionally high node degree centrality houses.

372 ***Figure 6: Hierarchy index. (a)*** *The frequency of occurrence of the node degree frequency for*
373 *a single participant. The green line indicates the linear regression starting at the median of the*
374 *distribution. **(b)** The distribution of the hierarchy index across all subjects.*

375

376   This observation triggers a detailed look at the mean node degree distribution over all
377 subjects to identify those special high node degree houses. Plotting the mean node degree of
378 houses averaged across subjects onto the map of Seahaven (Fig. 7c) highlights the scattered
379 subset of high node degree houses. Furthermore, the mean node degree across all houses and
380 all subjects measures 8.3 with a standard deviation of 3.98 (Fig. 7a). We select the value of
381 the 2-sigma distance (16.25) as the threshold to identify the high node degree houses. This
382 results in a set of 10 houses with node degree centrality values exceeding the threshold.
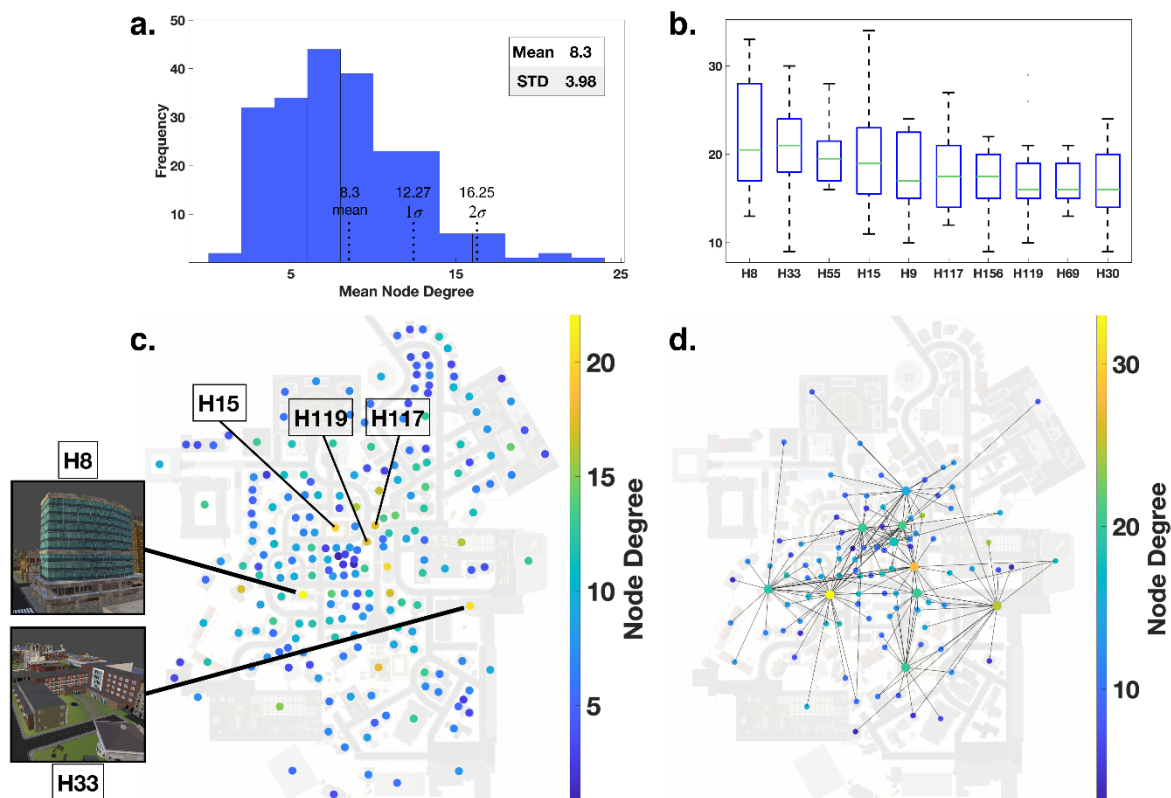
Figure 7: The high node degree centrality houses. (a) *The mean node degree distribution across all subjects with mean, 1σ- and 2σ-thresholds.* **(b)** *A box plot of the 10 houses with at least 2σ-distance to the mean node degree.* **(c)** *A map plot with all nodes color coded with their respective average node degree across all subjects.* **(d)** *The 10 houses, which exceeded the 2σ-distance to the mean displayed on the map for our example subject with all their connections and color coded with their respective node degree.*

Next, we analyze these houses with respect to their interconnectivity and visibility in the city. The distribution of the node degree centrality over subjects for each of the 10 houses reveals considerable variance (Fig. 7b). However, all 10 houses for all participants have a node degree centrality of 9 or above. Plotting them jointly with all their edges onto the map of Seahaven for our example subject shows that they are located centrally within the city. However, these houses are connected to the outer areas of the city and together cover nearly the whole city (Fig. 7d). All in all, these 10 high node degree houses, are located centrally in the city and their connections reach out into outer areas, i.e. they are viewed from much of the city area.

The characteristics of high visibility of a small number and similar use over participants would be expected in landmarks. Our results show that the node degree distribution is similar across subjects and that only a few houses have exceptionally high node degrees. This is supported by the hierarchical configuration of the network. The 10 houses with an average node degree distribution exceeding the a 2-sigma threshold are more centrally located and had viewing connections into the outer areas covering nearly the whole city. All things considered, these findings suggest the notion that this set of houses is exceptional across

407 multiple domains and displays the characteristics expected from landmarks. Therefore, in the
408 following we will refer to these 10 buildings as "gaze-graph-defined landmarks".

409 The connections between the gaze-graph-defined landmarks – The rich
410 club coefficient

411 In this section we investigate whether the gaze-graph-defined landmarks serve as the core of
412 a network that could be used for navigation in the city. For a quantitative investigation, we
413 applied the concept of the rich club coefficient to our gaze graphs. The rich club coefficient is
414 a frequently practiced graph-theoretical method in network theory and was initially applied
415 in internet network analyses (36). Yet, the rich club coefficient has also been transferred to
416 neuroscientific contexts. The approach has been used to map out both subcortical and
417 neocortical hub regions and to show that those regions with high linkages are also highly
418 connected between each other and, thus, indeed form a rich club (37). In this study, the rich
419 club coefficient allows to quantify in how far gaze-graph-defined landmarks are preferentially
420 connected to each other.

421 We calculated the connectivity between nodes with a specific degree value using

422
$$RC(k) = \frac{2E_{\geq k}}{N_{\geq k}(N_{\geq k} - 1)}$$

423 with k as the set node degree of the rich club, $E_{\geq k}$ as the number of edges between nodes
424 with degree larger or equal to k, and $N_{\geq k}$ as the number of nodes with degree larger or equal
425 to k. Thus, the rich club coefficient is the fraction of edges instantiated between nodes of
426 degree k or larger and the total number of edges possible between nodes the same degree or
427 larger.

428 For the interpretation of the rich club coefficient, we need a baseline. For that purpose,
429 we compare generated random graphs with similar statistics of node degrees and calculate
430 their rich club coefficient. For each subject, we generate 1000 random graphs with the same
431 number of nodes as the respective original graph. Of these we select the 10 graphs with the
432 most similar, in terms of the two-sample Kolmogorov Smirnov test, distributions of node
433 degrees to the original distribution. Subsequently, we divide the original rich club coefficient
434 by the 10 random coefficients respectively and averaged. Thus, a value above 1 indicates the
435 existence of a rich club with the respective node degree.

436 As a final step we investigate the rich club coefficient as a function of the threshold
437 node degree. Due to the strong hierarchy index, the number of nodes decreases drastically
438 with increasing node degree, resulting in increasing uncertainty. Therefore, we cut off the plot
439 at 1σ distance to the mean. With increasing threshold, the interconnectivity of the rich club
440 steadily increases and reaches 1.5 for a node degree of 13 or higher (Fig. 8a). This
441 demonstrates that these nodes are interconnected much more than expected by chance. We
442 calculate the rich club for each subject and specifically mark the top 10 houses with the highest
443 frequency of being part of the rich club across all subjects (Fig 8b). These houses had the
444 averaged highest interconnectivity and were the same houses that were identified as the top
445 10 node degree houses, i.e., gaze-graph-defined landmarks, earlier. This gives evidence for a
446 highly interconnected network of gaze-graph-defined landmarks in the city, a rich club.
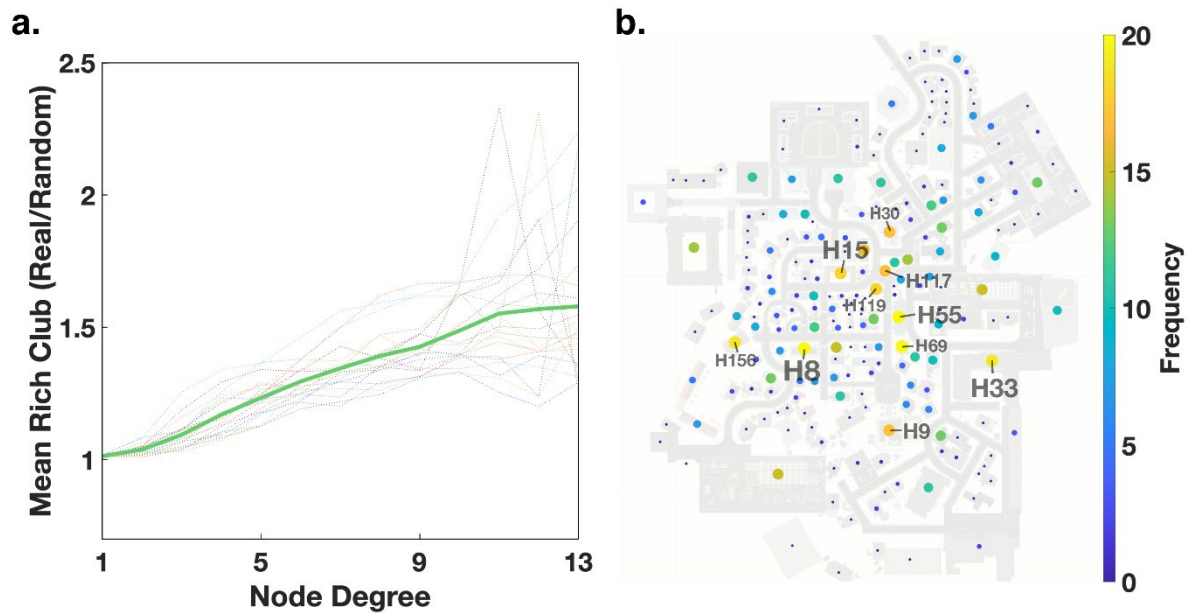
Figure 8: The rich club coefficient (a) *The development of the rich club coefficient with increasing node degree. The dot-lines are the rich club coefficients of individual participants, while the green line is the mean across all subjects.* **(b)** *All houses displayed on the map both color coded and size coded according to their frequency of being part of the rich club across subjects.*

Spatial arrangement of the gaze-graph-defined landmarks - Triangulation

To elucidate the role of the gaze-graph-defined landmarks in spatial navigation, we explore whether they could serve as a basis for triangulation. Triangulation is a method to infer the own location based on the viewing angle in respect to two location anchors. Our analysis has revealed that the gaze-graph-defined landmarks form a highly interconnected rich club. Thus, if the gaze-graph-defined landmarks are visible from most places in the city they could serve as a basis for triangulation.

As a first step we determine the parts of the city where at least one or two of the gaze-graph-defined landmarks were viewed by the subjects. We evaluate the spatial distribution of our participants during exploration of the virtual town and how many gaze-graph-defined landmarks were actually viewed from each location. This analysis is performed with a spatial resolution of 4x4m and an additional smoothing with a 3x3 unity kernel. Please note that this analysis depends on the actual gaze data and, thus, reflect from where participants actually did view the gaze-graph-defined landmarks. Next, we differentiated the locations in three categories: zero/one/two gaze-graph-defined landmark were viewed. The resulting map represents the potential of triangulation based on the gaze-graph-defined landmarks at different locations in the virtual city (Fig. 9). In 39.1% of the city areas that were visited by participants, two or more gaze-graph-defined landmarks were viewed from that location, providing the basis for triangulation. In an additional 32.7% of the city area, exactly one gaze-graph-defined landmark was viewed. Only in 28.1% of the visited city area, none of the gaze-graph-defined landmarks were viewed. Weighting the city areas with the absolute time

474   participants were located in, participants spend an even bigger fraction of the experiment
475   time in areas where the theoretical basis of triangulation was given. Specifically, participants
476   spent 53.2% of the experiment time in areas where at least two or more gaze-graph-defined
477   landmarks were viewed, 19.4% of the time in areas where one gaze-graph-defined landmark
478   was viewed and 27.4% of the time in areas where none was viewed. The latter regions were
479   mostly at the fringes of the city map. Overall, our results indicate that triangulation based on
480   gaze-graph-defined landmarks is possible in most parts of the city and participants spent the
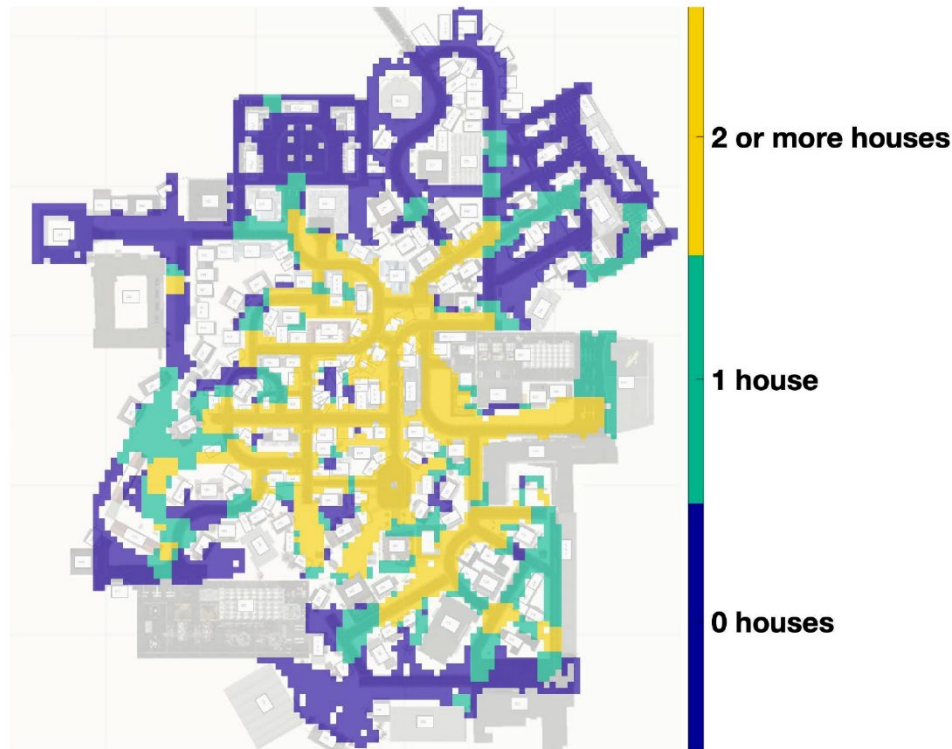481   majority of their time located in these areas.

482
483   **Figure 9: Triangulation**. *Location data of all participants plotted on the map of Seahaven.*
484   *The color code indicates how many of the gaze-graph-defined landmarks were viewed by*
485   *participants at each location.*

Discussion

487   In this study, we establish a method to quantify characteristics of visual behavior by using
488   graph-theoretical measures to abstract eye tracking data recorded in a 3D virtual urban
489   environment. We define gazes of subjects that freely explored the virtual city Seahaven, and
490   use these to convert the viewing behavior into graphs. In these gaze graphs, nodes represent
491   houses while edges represent their visual connection, i.e., gazes in direct succession on the
492   respective houses. Thus, the gaze graphs capture relevant spatial information gathered during
493   exploration. The node degree centrality graph measure reveals a surprisingly large variance.
494   However, the values of individual houses were rather consistent across subjects, as shown by
495   the high mean correlation. Additionally, we observed that the degree distribution across
496   houses increased steadily, except for only a few high node degree houses. These houses may
497   serve as import reference points, the so-called gaze-graph-defined landmarks. The analysis of

498 the hierarchy index demonstrates that the frequency of houses decreased drastically with
499 increasing node degree, revealing a hierarchical graph structure. The set of identified gaze-
500 graph-defined landmarks were indeed preferentially connected, as demonstrated by the rich
501 club coefficient. Finally, participants spent more than half their exploration time at locations,
502 where at least two of the houses of the rich club were viewed, allowing triangulation for
503 spatial localization. Thus, we presented a graph-theoretical approach for analyzing gaze
504 movements supporting spatial cognition.

505 In general, spatial navigation is based on multimodal sensory input. In the present
506 study we employ a virtual reality that is more restricted, i.e., limited to visual and vestibular
507 information. Recent literature highlights the importance of idiothetic and sensorimotor
508 information about self-position and about self-motion (38,39). Nevertheless, the dominating
509 sense of spatial perception is vision (40). Only this sense can gather reliable information of
510 space and the environment independently of the physical distance and also allows to perceive
511 topographic characteristics over large distances. Therefore, by observing visual attention and
512 visual behaviour using the method of eye tracking, we could gather important insights about
513 the usage of spatial cues, that indicate cognitive processes related to spatial navigation,
514 specifically landmark usage.

515 Basing our analysis on eye tracking data allows us to observe and investigate
516 participants visual attention during spatial navigation processing. However, the combination
517 of eye tracking and a head mounted virtual reality headset comes with several challenges.
518 Accuracy of mobile eye tracking systems if often reduced compared to other systems (41).
519 Specifically, in VR experiments, this is often due to the freedom of head movements and
520 weight of the VR headset. Since typical errors due to slippage and head movements increase
521 over time, we conducted a short validation and if necessary, a complete calibration validation
522 procedure every 5 minutes during the experiment (28). Nevertheless, the mean validation
523 error of 1.55° before experiment start and 1.88° during the experiment is rather high
524 compared to classical lab-based eye tracking studies. However, unlike lab-based eye tracking
525 studies, our preprocessing and analysis is based on hit point clusters that fell on the same
526 collider in the VR environment. Thus, we summarized data points that were located in rather
527 close proximity. Though the notion of close proximity must be taken with care since most
528 colliders still corresponded to the size of a complete house, therefore making a small deviation
529 in gaze location due to the validation error less problematic. All in all, considering that the
530 preprocessing is based on the spatial distribution of the accumulated hit point clusters, a
531 minor increase of validation error should not affect our data significantly.

532 One major factor in the analysis of eye tracking data recorded in an VR environment is
533 the algorithm differentiating the different types of eye movements and in our case the
534 definition and creation of the data form "gaze". Essentially, the four different types of eye
535 movements expected to occur in such a natural setting, i.e. saccade, fixation, vestibulo-occular
536 reflexes, and smooth pursuit movements, can be separated into two categories of visual input
537 perception. On the one hand, we have fixations as the typical source of visual perception.
538 Since vestibulo-occular reflexes and smooth pursuit movements stabilize the retinal image in
539 dynamic situations, they lead to visual perception similar to the input during fixations (29).

18

540   Thus, we can classify fixations, vestibulo-occular reflexes, and smooth pursuit movements as
541   the first category of visual perception. Saccades, on the other hand, render the participant
542   blind to the momentary visual input, hence a hit point sample created during a saccade will
543   not have been perceived by the participant (29,30). Consequently, it is essential to
544   differentiate the data between the first category of visual perception and saccades. Usually,
545   this is done either by velocity or gaze location-based algorithm. However, the virtual
546   environment in Unity3D results in three impeding factors. First, the complexity of the VR
547   environment in Unity3D only allowed for a 30 Hz sampling rate, thus it does not allow a
548   saccade detection based on sudden changes in gaze velocity. Secondly, the gaze location
549   calculated based on the ray cast process is limited to whole colliders, often covering the size
550   of a house, therefore it does not allow to identify small changes in gaze location. And thirdly,
551   depth perception is usually not accounted for in saccade detection algorithms, therefore
552   making it questionable to apply to eye tracking data recorded in a 3D. Consequently, the
553   available eye tracking data recorded in the VR environment did not provide the reliable
554   information necessary in classical fixation detection algorithms that allow to cleanly separate
555   the different eye movements. Therefore, defining the new data form "gaze" allows for a
556   functional method to clean the data. This process did not allow the identification of single eye
557   movements nor did it exclude all saccades from the data. However, by identifying the data
558   clusters that contained at least one fixation and excluded the data clusters that very likely did
559   not contain any fixation, we could clean the data from samples, that were unlikely to be
560   visually processed by the participant. This is further supported by our data, since the process
561   identified 86% of the data as gazes on average across all participants. With approximately 90%
562   of viewing time being expected to be spend on fixations (29), the process identifying gazes
563   appears to be on the conservative side. Overall, with higher sampling rates available in virtual
564   environments, new options to identify saccades might become available. However, we believe
565   that given the data we have available, defining the new data form "gaze" was the best option
566   to differentiate between stray samples unprocessed by the participant and meaningful data
567   carrying important information of the first category of visual perception including fixations,
568   vestibulo-occular reflexes, and smooth pursuit movements.
569   The process of spatial navigation is abundant in everyday life, therefore investigating
570   the process of gathering spatial navigation under natural conditions is our goal. While there
571   are some examples of studies in the field of spatial cognition conducted in natural
572   environments (42,43), these studies typically rely on behavioral data, and lack physiological
573   data recorded during the spatial navigation itself. With new mobile eye tracking systems
574   available, more studies investigating spatial cognition under natural conditions were
575   conducted in inside and outside environments. For example, Ohm et al. (2014) investigated
576   the selection process of landmarks in large scale indoor environments via the visual attraction
577   measured with mobile eye tracking (21) and evaluated pedestrian navigation systems based
578   on indoor landmarks (22). Kiefer et al. (2014) investigated self-localization based on
579   participants matching maps to the urban environment they were located at (20). Furthermore,
580   Wenczel et al. (2017) found differences in gaze behavior during incidental and intentional
581   route learning in a natural outdoor environment with intentional learning leading to more

582 fixations on landmarks (23). However, studies conducted with mobile eye tracking systems in
583 natural environments are usually challenged with several disadvantages. In addition to the
584 poor accuracy of mobile eye tracking systems, another major issue comes with identifying
585 frames with areas of interest during the data pre-processing. Often, identification of regions
586 of interests and therefore identifying the relevant fractions of the eye tracking data could only
587 be solved by manual detection in each frame (20,21,23). Furthermore, the natural
588 environment allows only limited control of the experimental conditions, especially regarding
589 variances in light affecting the eye tracker systems (20), and variance in the environment due
590 to other people or traffic (24,25). Depending on the system, eye tracking calibration can be
591 distorted for some distances, therefore limiting the valid distance of gazed objects that can be
592 analyzed (20). Consequently, most eye tracking studies struggle to be conducted under
593 natural conditions and thus reduce the ecological validity of gathered information of viewing
594 behavior under natural conditions. Instead, implementing experimental paradigms in VR
595 allows maximal control of experimental conditions while still providing a more naturalistic
596 environment and multisensory experience. Moreover, our analysis method using hit point
597 clusters and gaze events allow a fast and precise analysis option, therefore providing a
598 solution to the problem of detecting ROIs manually as observed in previous studies. Therefore,
599 new experimental paradigms in VR in combination with the eye tracking analysis proposed in
600 this paper provide a new option to investigate spatial navigation under naturalistic conditions.

601 The graph-theoretical measures used in this study is crucial, since it leads to a large
602 compression and abstraction of the data. Using this method, a few millions of gaze samples
603 were condensed into a few graphs. In general, graph theory is used in many areas to make
604 complex information of pairwise relations accessible (44–46). This process includes several
605 decisions, which might critically influence the later analysis. First, the visibility of the houses
606 depended on the participant's location in the virtual world, which is not necessarily close to
607 one of the respective houses. Accordingly, the participant might view two houses in direct
608 succession leading to a visual connection, even though the houses themselves might not be
609 visible from each other's location. Second, the gaze graph did not contain any information
610 about the order in which the two successive gaze events took place, consequently all gaze
611 graphs were undirected. Third, the graph contained binary, i.e., unweighted, edges.
612 Consequently, the gaze graphs only contained information about whether a visual connection
613 took place at some point during the experiment, and not how often. Fourth, in case the data
614 contained a cluster of missing data points no information about visual connections during this
615 time was available and no edge could be created between nodes. Applying the above
616 considerations for graph creation, our graphs represented the gaze data in a well-defined and
617 meaningful way.

618 Our results of the graph-theoretical analysis revealed a small subgroup of houses that
619 seemed to correspond to several characteristics we expected landmarks to have in a gaze
620 graph. Landmark knowledge refers to the knowledge of salient objects that serve as
621 orientation anchors and are memorized when exploring a new environment. (2,4,5). Since the
622 node degree is a common measure to investigate the importance of single nodes in the
623 network (47), we expected landmarks to stand out in visual behavior compared to other

624 houses and therefore show high importance in the gaze graph visible in high values of their
625 node degree centrality. The gaze graph represented the visual connections between houses,
626 therefore, the node degree centrality measured how many different houses were visually
627 connected with the house in question. If a house would serve as a landmark, we would expect
628 participants to often create visual connections between the landmark house and other houses
629 while trying to navigate, especially when they try to relate newly learned houses in respect to
630 the landmark house. The results of our node degree centrality analysis showed a clear
631 difference between the subgroup of the rich club compared to the other houses. Moreover,
632 the average node degree values of the houses exceeded the mean node degree over two
633 times the standard deviation. The rarity of these buildings is strengthened by the high mean
634 hierarchy index. Taking all these findings into account, we decided to call this subgroup of
635 houses gaze-graph-defined landmarks, since they fit the characteristics a landmark would
636 display in a gaze graph.

637 While it is undisputed that landmarks are important for spatial navigation, the details
638 about their functionality are not yet completely understood. Commonly, it is assumed that
639 landmarks form the basic building blocks of landmark knowledge that is then extended to
640 route and survey knowledge by gradually connecting landmarks with routes and achieving
641 knowledge about the relational information of the landmark locations (2). Others have
642 proposed a more continuous and parallel development of landmark, route and survey
643 knowledge (4,42). Nonetheless, especially the mechanisms of how landmarks relate to mental
644 maps, how landmark knowledge is integrated into survey knowledge or which functions
645 landmarks maintain when a map-like survey knowledge is already available remain unclear.
646 By calculating the rich club coefficient of all gaze graphs, we found a causality between
647 increasing node degree and increasing connectivity between the respective nodes. Gaze
648 graph-defined landmarks were above chance level interconnected to each other. Moreover,
649 the interconnected gaze-graph-defined landmarks of the city seem to form a network, i.e. a
650 rich club, covering a large, mostly central part of the city. This could be a first indication that
651 landmarks are not only used as orientation anchors, but could form an underlying network of
652 orientation anchors that span out the framework of a mental map. Consequently, landmarks
653 could anchor the mental map and thus, serve as an important feature of survey knowledge.

654 Furthermore, our results revealed that participants spent a large fraction of their time
655 in the areas where the theoretical basis of triangulation with gaze-graph-defined landmarks
656 was given. Specifically, participants spend more than 50% or their experiment time in the
657 locations where at least two of the gaze-graph-defined landmarks were visible. Interestingly,
658 these areas were located at the more central regions of the city. This could be explained by
659 several reasons. On the one hand, participants could prefer city areas in which they could
660 triangulate based on gaze-graph-defined landmarks and consequently spend most of their
661 time in these locations. On the other hand, this could also be related to the size of Seahaven.
662 It was shown that within 90 min total exploration time participants explored all areas of
663 Seahaven, central areas a slightly more often than peripheral areas (48). Furthermore, our
664 results indicate that participants created a network of landmarks. This suggests that the
665 subjects deliberately searched for houses that could serve as landmarks. The selection of 10

666   gaze-graph-defined landmarks constituted to less than 5% of all available houses located in
667   Seahaven. Consequently, our results strongly indicate that participants were not only looking
668   for landmarks, but also quite strategically chose landmarks as orientation anchors that
669   provided a maximum amount of information on the basis of the rather small subset of less
670   than 5% of all city buildings. Therefore, our results seem to support the notion that landmarks
671   were not only selected based on saliency but seemed to follow a specific strategy maximizing
672   the navigational benefits while minimizing processing expenses.

673   This interpretation of our results was supported by further characteristics of the gaze-
674   graph-defined landmarks. In a navigational context, the term landmark usually refers to any
675   type of object that is highly visible or easily recognizable in the environment and thus serves
676   as a point of reference (49), although the overall saliency also depends on "a unique property
677   of the trilateral relation between the feature itself, the surrounding environment and the
678   observer's point of view" (50). In addition, landmarks are often differentiated into local and
679   global landmarks. Typically, global landmarks can be seen from far distances and provide a
680   reference for directional and more compass-like orientation, whereas local landmarks might
681   only be visible in a local area and are often located at road crossings (51). Looking at the visual
682   appearance of the gaze-graph-defined landmarks, we found typical characteristics of
683   landmarks including visual saliency due to size, color and location in the city. For example, the
684   house with the highest average node degree across all participants is higher than most
685   surrounding buildings and has a large distinct blue window front that sets it aside from its
686   surrounding (Fig. 6b,7b). Additionally, it is located in the very center of the city making it visible
687   from most parts of the city. The house with the second highest average node degree across
688   all participants stood out due to its size regarding its surface area and location next to the
689   main road that connects the most north and most south part of the city, even though it was
690   only visibly in the south-east part of the city. In general, most of the gaze-graph-defined
691   landmarks were located next to crossings of main roads in the city thus fitting a characteristic
692   of local landmarks (51). Our results revealed that the gaze-graph-defined landmarks also had
693   features of global and local landmarks.

694   While the application of graph theory has enabled us to use a variety of already
695   established graph-theoretical measures to analyze the gaze graphs and resulted in very
696   promising results, only a small amount of the available graph-theoretical measures was
697   applied during our analysis. For example, the node degree centrality is defined as the sum of
698   connections of each node. Generally speaking, nodes with a lot of connections are likely to be
699   important for most networks (45) which is why the node degree centrality is widely used in
700   graph-theoretical analyses and serves as the basis of our gaze graph analysis (52). According
701   to Sporns (2018), the usage of graph theory in neuroscientific studies has increased in the
702   recent years and the node degree centrality can be a useful measure for network analyses.
703   However, these analyses are mostly based on pairwise dyadic approaches and the full
704   potential of graph theory has not yet been applied.

705   A variety of graph measures is available for analyzing networks and node importance
706   can be defined in different ways, exceeding the node degree-based analysis. A particularly
707   important measure in social network analyses is the so-called betweenness centrality (53).

708   The measure counts how many shortest paths cross a particular node that is "between" a lot
709   of nodes. In social network analyses, betweenness centrality is beginning to replace the node
710   degree centrality to explain social network dynamics with respect to the importance of nodes
711   with high node betweenness for attracting and strengthening new links (54). Consequently,
712   the betweenness centrality could be a potential candidate for explaining spatial knowledge
713   acquisition. As mentioned, the centrality measure gains importance within the field of social
714   network analysis and is used to characterize the attraction and strengthening of connections.
715   A person, represented by a node in a social network, would have a high betweenness
716   centrality if the person connects a variety of other persons that themselves do not know the
717   connecting person. By transferring this thought to spatial navigation, a house with high
718   betweenness centrality would connect the views to two buildings that are not viewed in direct
719   succession themselves. Thus, this building serves as an anchor point for these two buildings
720   and forms a part of the route between them. Betweenness centrality could serve as a measure
721   for characterizing the gathered route knowledge. Overall, graph theory offers a variety of
722   different measures and has not yet reached its full potential within neuroscientific research.

723   Overall, our results establish a new methodology to process eye tracking data in
724   complex 3D environments and identify and assess the function of landmarks in spatial
725   navigation. Applying this methodology provides a new and unique insight into behavioral data
726   of visual attention processes during spatial navigation and opens the door for a novel
727   approach to investigate spatial navigation. To fully unlock the potential of graph theory, we
728   propose additional graph-theoretical measures to investigate gaze graphs in the future.
729   Specifically, we consider the betweenness centrality that could help to understand the
730   formation of spatial knowledge beyond landmark knowledge.

Methods

731

The virtual town of Seahaven

732

733   The virtual town of Seahaven was built to investigate spatial learning during free exploration
734   (28,48,55). In total, Seahaven contains 213 houses in varying size and shape. Furthermore, the
735   city was designed as a connected urban space of roughly 216,000 m2 (48). The street structure
736   consisted of winding, small and big roads and overall resembled a European city center. The
737   entire virtual environment reflects natural spatial relations, with one Unity unit corresponding
738   to one meter. The virtual height of the participant was set to 2 meters and was unified for
739   every participant. To provide a better frame rate in Unity, a far clipping plane in a distance of
740   160 meters was introduced. Consequently, no objects were visible to the participants that
741   were located further than 160m away from current participant location.

Structure of the experiment

742

743   In total, 22 participants performed the complete experiment. This included three sessions
744   within at most 10 days. Each session consisted of 5 parts: (1) a brief introduction to the
745   experiment; (2) examples of the spatial tasks to be completed after the last recording session;

746 (3) preparation of the VR setup, including the adjustment of VR-headset with eye tracker, the
747 calibration and validation of the eye tracker, and in the 1st session a movement training on the
748 virtual island; (4) the main experimental phase, i.e. exploring the virtual town for 30 minutes
749 while movements and eye tracking data were recorded; (5) three spatial tasks performed
750 outside of the VR, which are covered in detail in (48). These data are, however, are outside
751 the scope of the present paper and not further covered here.

Laboratory setup

753 Participants wore a head mounted HTC Vive virtual reality headset and were seated on a
754 swivel chair. To prevent limitations on rotations as well as removing the tactile directional
755 feedback due to hanging cables, a vertical cable solution was implemented. Participants
756 moved using the HTC controller at walking speed. To decrease the risk of motion sickness,
757 participants were instructed to only walk forward with the controller. If they wanted to switch
758 directions or turn, they were instructed to stop walking and then rotate their entire body with
759 the chair in the desired direction.

Eye tracking

761 A pupil labs eye tracker was directly integrated in the HTC Vive headset (refresh rate 120Hz,
762 gaze accuracy 1,0°, gaze precision 0.08°, visual field 110°) (28). Both calibration and validation
763 were executed in the virtual reality while the participants where still on a separate training
764 island. For each subject, a 17-point calibration and a 9-point validation were conducted until
765 the validation error was below 2° (on average 1.53°). During the experiment, a 1-point
766 validation was performed every 5 minutes ensuring the correct tracking precision. If the error
767 exceeded 2°, a complete 17-point calibration and 9-point validation was performed until the
768 error was below the original threshold of 2° (mean 3.14°, median 2.11° before and mean 1.88°,
769 median 1.28° after a new calibration). 55% of all 1-point validations had a validation error
770 above 2° and had to be recalibrated, hence, highlighting the importance of regular validation
771 control. When the calibration and validation process was performed during the exploration of
772 Seahaven, the display of the virtual city disappeared until the validation was completed.

Acknowledgments

Conflict of Interest Statement

778 The authors have declared that no competing interests exist.

Funding

Data Availability

All data files the described pre-processing and analysis are based on, are available at the Center for Open Science https://osf.io/aurjk/, DOI 10.17605/OSF.IO/AURJK. All pre-processing, visualization and analysis scripts supporting this publication are available on our Github repository including an extensive documentation to allow full reproducibility:
https://github.com/JasminLWalter/FindingLandmarks_a_publication_repository

References

1.   Kelly JW, McNamara TP. Spatial Memory and Spatial Orientation. In: Freksa C, Newcombe NS, Gärdenfors P, Wölfl S, editors. Spatial Cognition VI Learning, Reasoning, and Talking about Space. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 22–38.

2.   Siegel AW, White SH. The Development of Spatial Representations of Large-Scale Environments. Adv Child Dev Behav. 10:9–55. doi:10.1016/S0065-2407(08)60007-5

3.   Lynch K. The image of the city. Vol. 11, MIT press.

4.   Montello DR. A new framework for understanding the acquisition of spatial knowledge in large-scale environments. Spat temporal Reason Geogr Inf Syst. :143–54.

5.   Klippel A, Winter S. Structural Salience of Landmarks for Route Directions. In: Cohn AG, Mark DM, editors. Spatial Information Theory. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 347–62. doi:10.1007/11556114_22

6.   Meilinger T, Frankenstein J, Bülthoff HH. Learning to navigate: Experience versus maps. Cognition. 129(1):24–30. doi:10.1016/J.COGNITION.2013.05.013

7.   O'Keefe J, Nadel L. The hippocampus as a cognitive map. Oxford Univ Press.

8.   Sholl MJ. Cognitive Maps as Orienting Schemata. J Exp Psychol Learn Mem Cogn. 13(4):615–28. doi:10.1037/0278-7393.13.4.615

9.   Thorndyke PW, Hayes-Roth B. Differences in spatial knowledge acquired from maps and navigation. Cogn Psychol. 14(4):560–89. doi:10.1016/0010-0285(82)90019-6

10. Morris RGM. Spatial localization does not require the presence of local cues. Learn Motiv. 12(2):239–60. doi:10.1016/0023-9690(81)90020-5

809   11.   McDonald RJ, White NM. Parallel information processing in the water maze: Evidence
810   for independent memory systems involving dorsal striatum and hippocampus. Behav
811   Neural Biol. 61(3):260–70. doi:10.1016/S0163-1047(05)80009-3

812   12.   Packard MG, McGaugh JL. Inactivation of hippocampus or caudate nucleus with
813   lidocaine differentially affects expression of place and response learning. Neurobiol
814   Learn Mem. 65(1):65–72. doi:10.1006/NLME.1996.0007

815   13.   Mueller SC, Jackson CPT, Skelton RW. Sex differences in a virtual water maze: An eye
816   tracking   and   pupillometry   study.   Behav   Brain   Res.   193(2):209–15.
817   doi:10.1016/j.bbr.2008.05.017

818   14.   Iaria G, Petrides M, Dagher A, Pike B, Bohbot VD. Cognitive strategies dependent on the
819   hippocampus and caudate nucleus in human navigation: Variability and change with
820   practice. J Neurosci. 23(13):5945–52. doi:10.1523/jneurosci.23-13-05945.2003

821   15.   Hamilton DA, Driscoll I, Sutherland RJ. Human place learning in a virtual Morris water
822   task: Some important constraints on the flexibility of place navigation. Behav Brain Res.
823   129(1–2):159–70. doi:10.1016/S0166-4328(01)00343-6

824   16.   Andersen NE, Dahmani L, Konishi K, Bohbot VD. Eye tracking, strategies, and sex
825   differences   in   virtual   navigation.   Neurobiol   Learn   Mem.   97(1):81–9.
826   doi:10.1016/j.nlm.2011.09.007

827   17.   Newman EL, Caplan JB, Kirschen MP, Korolev IO, Sekuler R, Kahana MJ. Learning your
828   way around town: How virtual taxicab drivers learn to use both layout and landmark
829   information. Cognition. 104:231–53. doi:10.1016/j.cognition.2006.05.013

830   18.   Farran EK, Formby S, Daniyal F, Holmes T, Van Herwegen J. Route-learning strategies in
831   typical and atypical development; eye tracking reveals atypical landmark selection in
832   Williams syndrome. J Intellect Disabil Res. 60(10):933–44. doi:10.1111/jir.12331

833   19.   Hejtmánek L, Oravcová I, Motýl J, Horáček J, Fajnerová I. Spatial knowledge impairment
834   after GPS guided navigation: Eye-tracking study in a virtual town. Int J Hum Comput
835   Stud. 116(May 2017):15–24. doi:10.1016/j.ijhcs.2018.04.006

836   20.   Kiefer P, Giannopoulos I, Raubal M. Where am i? Investigating map matching during
837   self-localization with mobile eye tracking in an urban environment. Trans GIS.
838   18(5):660–86. doi:10.1111/tgis.12067

839   21.   Ohm C, Müller M, Ludwig B, Bienk S. Where is the Landmark? Eye tracking studies in
840   large-scale indoor environments. CEUR Workshop Proc. 1241:47–51.

841   22.   Ohm C, Müller M, Ludwig B. Evaluating indoor pedestrian navigation interfaces using
842   mobile   eye   tracking.   Spat   Cogn   Comput.   17(1–2):89–120.
843   doi:10.1080/13875868.2016.1219913

26

844 23. Wenczel F, Hepperle L, von Stülpnagel R. Gaze behavior during incidental and
845   intentional navigation in an outdoor environment. Spat Cogn Comput. 17(1–2):121–42.
846   doi:10.1080/13875868.2016.1226838

847 24. Kiefer P, Straub F, Raubal M. Location-Aware Mobile Eye Tracking for the Explanation
848   of Wayfinding Behavior. In: Gensel J, Josselin D, Vandenbroucke D, editors. Proceedings
849   of the AGILE'2012 International Conference on Geographic Information Science.

850 25. Kiefer P, Straub F, Raubal M. Towards location-aware mobile eye tracking. Eye Track
851   Res Appl Symp. :313–6. doi:10.1145/2168556.2168624

852 26. Kim JY, Kim MJ. Exploring visual perceptions of spatial information for wayfinding in
853   virtual reality environments. Appl Sci. 10(10):1–15. doi:10.3390/app10103461

854 27. Meilinger T, Frankenstein J, Watanabe K, Bülthoff HH, Hölscher C. Reference frames in
855   learning from maps and navigation. Psychol Res. 79(6):1000–8. doi:10.1007/s00426-
856   014-0629-6

857 28. Clay V, König P, König SU. Eye tracking in virtual reality. J Eye Mov Res. 12(1).
858   doi:10.16910/jemr.12.1.3

859 29. Duchowski AT. Eye Tracking Methodology [Internet]. Third Edit. Eye Tracking
860   Methodology. Springer; doi:10.1007/978-3-319-57883-5

861 30. Breitmeyer BG, Ganz L. Implications of sustained and transient channels for theories of
862   visual pattern masking, saccadic suppression, and information processing. Psychol Rev.
863   83(1):1–36. doi:10.1037/0033-295X.83.1.1

864 31. Reichle ED, Pollatsek A, Fisher DL, Rayner K. Toward a Model of Eye Movement Control
865   in Reading. Psychol Rev. 105(1):125–57. doi:10.1037/0033-295X.105.1.125

866 32. Nuthmann A, Smith TJ, Engbert R, Henderson JM. CRISP: A Computational Model of
867   Fixation   Durations   in   Scene   Viewing.   Psychol   Rev.   117(2):382–405.
868   doi:10.1037/a0018924

869 33. Fiedler M. Algebraic connectivity of graphs. Czechoslov Math J. 23(2):298–305.
870   doi:10.21136/CMJ.1973.101168

871 34. Fiedler M. Laplacian of graphs and algebraic connectivity. Vol. 25, Banach Center
872   Publications. p. 57–70. doi:10.4064/-25-1-57-70

873 35. Rodrigue J-P. The Geography of Transport Systems. 5th ed. Routledge; 480 p.

874 36. Zhou S, Mondragón RJ. The rich-club phenomenon in the internet topology. IEEE
875   Commun Lett. 8(3):180–2. doi:10.1109/LCOMM.2004.823426

876   37.   Van Den Heuvel MP, Sporns O. Rich-club organization of the human connectome. J Neurosci. 31(44):15775–86. doi:10.1523/JNEUROSCI.3539-11.2011

878   38.   Colombo D, Serino S, Tuena C, Pedroli E, Dakanalis A, Cipresso P, et al. Egocentric and allocentric spatial reference frames in aging: A systematic review [Internet]. Vol. 80, Neuroscience and Biobehavioral Reviews. Pergamon; p. 605–21. doi:10.1016/j.neubiorev.2017.07.012

882   39.   Chrastil ER, Warren WH. Active and passive contributions to spatial learning. Psychon Bull Rev. 19(1):1–23. doi:10.3758/s13423-011-0182-x

884   40.   Ekstrom AD. Why vision is important to how we navigate. Hippocampus. 25(6):731–5. doi:10.1002/hipo.22449

886   41.   Ehinger B V., Groß K, Ibs I, König P. A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. PeerJ. 2019(7):1–43. doi:10.7717/peerj.7086

889   42.   Ishikawa T, Montello DR. Spatial knowledge acquisition from direct experience in the environment: Individual diVerences in the development of metric knowledge and the integration of separately learned placesIshikawa ¤, T., & Montello, D. R. (2006). Spatial knowledge acquisition f. Cogn Psychol. 52(2):93–129. doi:10.1016/j.cogpsych.2005.08.003

894   43.   Ishikawa T, Fujiwara H, Imai O, Okabe A. Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. J Environ Psychol. 28(1):74–82. doi:10.1016/J.JENVP.2007.09.002

897   44.   Mondal B, De K. Overview Applications of Graph Theory in Real Field. Int J Sci Res Comput Sci Eng Inf Technol. 2(5):751–9.

899   45.   Sporns O. Graph theory methods: applications in brain networks. Dialogues Clin Neurosci. 20(2):111–20. doi:10.31887/DCNS.2018.20.2/osporns

901   46.   Singh RP. Application of Graph Theory in Computer Science and Engineering. Int J Comput Appl. 104(1):10–3. doi:10.5120/18165-9025

903   47.   Srinivasan S, Hyman JD, O'Malley D, Karra S, Viswanathan HS, Srinivasan G. Chapter Three - Machine learning techniques for fractured media. Moseley B, Krischer L, editors. Vol. 61, Advances in Geophysics. Elsevier; 109–150 p. doi:10.1016/BS.AGPH.2020.08.001

907   48.   König SU, Keshava A, Clay V, Rittershofer K, Kuske N, König P. Embodied Spatial Knowledge Acquisition in Immersive Virtual Reality: Comparison to Map Exploration. Front Virtual Real. 2. doi:10.3389/frvir.2021.625548

910   49.   Presson CC, Montello DR. Points of reference in spatial cognition: Stalking the elusive
911   landmark*. Br J Dev Psychol. 6(4):378–81. doi:10.1111/j.2044-835x.1988.tb01113.x

912   50.   Caduff D, Timpf S. On the assessment of landmark salience for human navigation. Cogn
913   Process. 9(4):249–67. doi:10.1007/s10339-007-0199-2

914   51.   Steck SD, Mallot HA. The role of global and local landmarks in virtual environment
915   navigation. Presence Teleoperators Virtual Environ. 9(1):69–83.
916   doi:10.1162/105474600566628

917   52.   Farahani F V., Karwowski W, Lighthall NR. Application of graph theory for identifying
918   connectivity patterns in human brain networks: A systematic review. Front Neurosci.
919   13(JUN):1–27. doi:10.3389/fnins.2019.00585

920   53.   Diestel R. Graph Theory [Internet]. Fifth Edit. Axler S, Ribet K, editors. Graduate Texts
921   in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg; 49–64 p. (Graduate
922   Texts in Mathematics; vol. 173). doi:10.1007/978-3-662-53622-3

923   54.   Topirceanu A, Udrescu M, Marculescu R. Weighted Betweenness Preferential
924   Attachment: A New Mechanism Explaining Social Network Formation and Evolution. Sci
925   Rep. 8(1):1–14. doi:10.1038/s41598-018-29224-w

926   55.   König SU, Clay V, Nolte D, Duesberg L, Kuske N, König P. Learning of Spatial Properties
927   of a Large-Scale Virtual City With an Interactive Map. Front Hum Neurosci. 13:240.
928   doi:10.3389/FNHUM.2019.00240