Pointing Errors in Non-Metric Virtual Environments

Alexander Muryy and Andrew Glennerster[1]

School of Psychology and Clinical Language Sciences,
University of Reading, Reading, RG6 6AL, UK
[1]a.glennerster@reading.ac.uk

Abstract. There have been suggestions that human navigation may depend on representations that have no metric, Euclidean interpretation but that hypothesis remains contentious. An alternative is that observers build a consistent 3D representation of space. Using immersive virtual reality, we measured the ability of observers to point to targets in mazes that had zero, one or three 'wormholes' – regions where the maze changed in configuration (invisibly). In one model, we allowed the configuration of the maze to vary to best explain the pointing data; in a second model we also allowed the local reference frame to be rotated through 90, 180 or 270 degrees. The latter model outperformed the former in the wormhole conditions, inconsistent with a Euclidean cognitive map.

Keywords: Human Navigation, Spatial Representation, Virtual Reality, Metric model, Motion Parallax, Binocular Disparity, Topological Model, Labelled Graph, View-Based.

1 Introduction

During active exploration of a 3D scene, an observer must build up some kind of representation about the layout of objects that will be useful from a different vantage point. One hypothesis is that the representation corresponds to a type of map with metric measures such as distances and angles conforming to a Euclidean geometry, even if these are not a faithful reproduction of the environment [1–3]. Such a metric cognitive map could be constructed, for instance, on the basis of path integration. It could provide a comprehensive description of the environment and underpin a wide variety of spatial tasks such as general navigation, finding shortcuts or detours and pointing to targets. Crucially, the structure of the data stored in such a map would be independent of the task. However, while there is a predominant view that metric representations may provide an adequate description of small open environments (vista spaces), there is good evidence that this hypothesis does not hold for a global representation of large complex environments [4]. Multiple experimental studies speak against global metric representations [4–9]. For instance, Warren et al [5] found that perceived locations of targets in a labyrinth may overlap, i.e. the perceived geometry of space is not metrically consistent (they report discontinuities in spatial representation, such as "rips" and "folds"). Typically, human participants can navigate efficiently through complicated

experimental environments which supports the notion of a global spatial representation. However, their perception of metric qualities is often distorted. For instance, perceived length of a route depends on the number of turns and junctions it contains [9, 10]. Angular and directional judgments are highly unreliable [7, 11–13] and perceived angles between junctions are biased towards 90° [9, 14].

An alternative representation of the environment that does not require metric measures and coordinate frame is a topological one [5, 6, 8, 9, 15–17]. A topological representation consists of a network of nodes and links that connect them. The exact shape of the path between nodes is not defined because the edges simply represent the connectivity of the environment, e.g. the existence of the path between two locations. Nevertheless, topological knowledge of an environment is sufficient for general navigation and allows the observer to find shortcuts and detours. This type of representation has been used to account for human navigational behaviour and has been applied in robot navigation systems [16, 18].

Hybrid models, that include both topological information of space connectivity and metric information have also been proposed [5, 8, 15, 19, 20]. Metric information about scene layout is often assumed to be available in the region that is visible from one location, e.g. within a room. Meilinger has described these as 'vista spaces' [19]. Such representations are hierarchical in nature, such that metric representation is reliable locally at the level of vista spaces but at the same time the global metric representation may be distorted in both the location of objects and the orientation of local 'vista spaces'. Evidence compatible with representations of this type has been reported by [4, 5, 9] on the basis of pointing, walking or other orientation judgements in environments that were known to the participant. Another aspect of spatial behaviour that is not predicted by a metric map is that the retrieval of information may depend on the observer's location [14], so that perception of the spatial layout of the scene is different when judged from point A and point B.

In our experiments, we generated virtual environments that were impossible to recreate physically and tested participants' ability to point to targets that they had encountered previously but could not currently see. The reason that such environments are informative is that they allow predictions to be tested that could not be distinguished in a normal environment. Non-Euclidean environments[AG1] of this sort were used previously to test human cognitive maps, for example [5, 21–23]. Warren and Chrastil [5] created a virtual labyrinth with "wormholes" that teleported participants smoothly between locations. Vasylevska and Kaufmann [22] tested environments with spatially overlapping regions much like the environments in our experiments in order to simulate space compression for VR applications. Zetzsche [23] and Klus [21] developed impossible virtual environments that violate Euclidean metrics and planar topology. Surprisingly, in all these studies human participants showed remarkable insensitivity to metric inconsistencies of space. Warren [5] used 'as-the-crow-flies' walking as a measure of the perceived distance and direction of a previously-seen target. Warren et al concluded that participants were using a distorted type of map, a 'labelled graph', in this task. The goal of our experiment was similar but, in our case, we explored a wider range of metric configurations that might explain the pointing data and a wider range of non-metric configurations, in which local reference frames were allowed to have different orientations.

2 Methods

2.1 Participants

The 8 participants (3 male and 5 female) were students or members of the School of Psychology and Clinical Language Sciences. All participants had normal or corrected to normal vision, one participant wore glasses during experiment, and all had good stereo-acuity (TNO stereo test, 60 arcsec or better). All participants were naïve to the purpose of the study. Participants were given a one hour practice session in VR to familiarize them with our set-up in a simplified virtual environment (open room with targets in boxes, but no inner walls) and metric versions of the mazes. Those who experienced motion sickness during the practice session or who preferred not to continue stopped at this stage. Altogether there were 7 sessions (including practice) roughly 1 hour each, all sessions were conducted on different days. Participants were advised not to stay in VR longer than 10 minutes and were reminded to take breaks. They received a reward of 12 pounds per hour. The study received approval of the research ethic committee of the university of Reading.

2.2 Experimental set-up

The experiment took place in the VR laboratory equipped with Vicon tracking system (14 infrared cameras) that provides 6 d.o.f. information about the headset position and orientation with nominal accuracy ±0.1 mm and 0.15° respectively. We used nVisor SX111 head mounted display with 111° field of view and binocular overlap of 50°. The position and orientation of the HMD was tracked at 240Hz and passed to the graphics PC with a GTX 1080 video card. The stimuli were designed using Unity 3D system and rendered online at 60fps. Participants were allowed to walk freely and explore the virtual environment in a natural way. The virtual environment occupied a physical space of 3x3m in the lab. The virtual labyrinth was originally a 5 by 5m environment with corridors in the maze 1m wide. In order to fit in the 3 by 3m lab, the labyrinth was shrunk to 0.6 scale (e.g. 60cm wide corridors) and displayed so that the floor was 1m below eye height for all participants Participants generally found this acceptable and did not notice that the room was not normal size, compatible with previous experiments [24]. During the experiment, the participants wore a virtual wristband that provided information about the task (shown, for illustrative purposes, at the top of Fig. 1B). Participants were also equipped with a pointing device that they used during the pointing phase.

2.3 Stimuli

We designed 2 general layouts of the virtual labyrinth. The physical size of both labyrinths was 3x3m and the width of the corridors was 60cm, see Fig. 1a. Thus, the virtual environment could be subdivided into 25 (5x5) elementary squares of the size of corridor's width. The green cylinder (Fig. 1a) indicates the starting location. In order

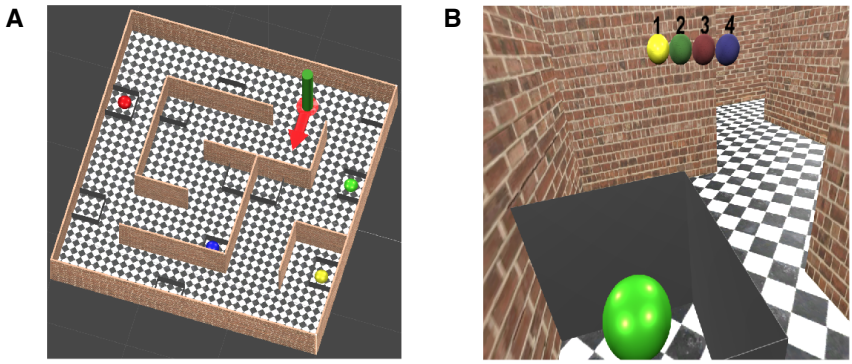

Fig. 1. Views of the labyrinth. A) View from above. B) First person view. The green target is visible inside a grey box. The target sequence is shown on top (YGRB) and the current target is highlighted (yellow). In the real experiment, this information was displayed on a virtual wristband on the participant's non-dominant hand.

to start the experiment, the participant should stand at the start location and look along the corridor's width. The green cylinder (Fig. 1a) indicates the starting location. In order to start the experiment, the participant should stand at the start location and look along the direction indicated by the red arrow. At the beginning of the experiment the green cylinder and red arrow disappear, so that starting location is not marked during the exploration phase. The labyrinth contains 4 target objects (red, green, blue and yellow spheres) hidden inside open gray boxes, so that they could be seen only from a short distance. A number of empty grey boxes were also added as distractors. Fig. 1b shows a first-person view along a virtual corridor containing the green target inside a grey box. Also, the participants wore a virtual wristband that displayed information about the currents task, such as the sequence in which targets should be collected (e.g. Yellow-Green-Red-Blue, see Fig. 1b, top).

For each labyrinth, we varied the complexity of the environment gradually by extending the length of the corridors with non-metric wormholes, see Fig. 2b and c. There were 3 conditions per scene: one metric (Fig. 2a) and two non-metric containing 1 and 3 wormhole correspondingly. Start is marked as 'S', colored circles indicate targets. The wormholes are indicated with letter W surrounded by colored dashed lines in the schematics (Fig. 2). The dashed lines act as invisible triggers, when participant crosses the triggers the environment changes as shown in the sub-plots. For instance, in 1-wormhole condition shown in Fig. 2b as participant crosses the red trigger the environment changes to schematic W1(a), as the participant continues walking down the path through the wormhole (affordable walking area inside a wormhole is marked by the dashed black lined) and crosses the green trigger, the environment changes again to schematic W1(b), then as the participant crosses the blue trigger he or she exits the wormhole and the environment changes back to the general layout. Participants were not aware of the presence of the triggers because the local views of the environment did not change during the trigger crossing.

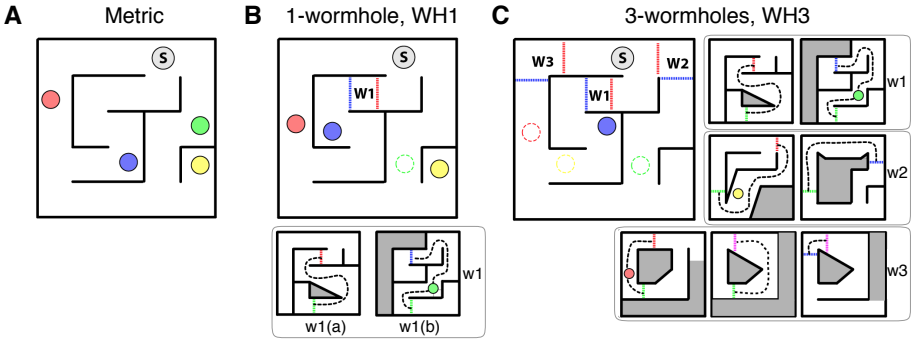

Fig. 2. Schematics of the labyrinth. A) Metric condition. B) Non-metric 1-wormhole condition. The green target was placed inside the wormhole. 3) Non-metric 3-wormholes condition. The general layout (containing Start, which is marked as 'S') remained constant between conditions. The wormholes are marked with letters W surrounded by red and blue triggers. As the participant crossed a trigger, the environment changed according to the sub-schematics. Inside a wormhole, the participant could only walk along the route marked by black dashed line. These schematics show Scene 1. Equivalent set of maps were generated for Scene 2 (see layout in Fig. 10).

For Scene 1, the same original or 'base level' layout applied in all three conditions (metric, one-wormhole and 3-wormholes), as shown in Fig. 2a. Similar principles applied to Scene 2 (which also had a metric, one-wormhole and three-wormhole version). The corridors through the wormholes did not have junctions, thus the topological connectivity of space was the same in all 3 conditions. The only difference between metric and non-metric conditions is the length and the shape of the corridors. The wormholes extend the corridors in a non-metric way and make a global metric representation impossible. For instance, the path through the wormhole in Fig. 2b has the shape of the figure of eight, i.e. it crosses itself although there are no junctions along that path, which is physically impossible. Notice also that the wormhole overlaps widely with the 'base level' layout.

2.4 Procedure

Participants were asked to perform a navigational and pointing tasks in a virtual labyrinth (see Fig. 1). During a single experimental session that took about 1 hour we used labyrinths with the same general layout in 3 conditions: metric, non-metric with 1 wormhole and non-metric with 3 wormholes. The three conditions were tested sequentially and participants could take breaks in between. Participants always started with the metric condition, followed by the 1-wormhole and then the 3-wormhole conditions, thus increasing the complexity gradually. The tasks and instructions were identical for all 3 conditions.

For each condition participants had to complete 8 rounds of walking and pointing. Each round started with a navigational task during which the participants were asked to collect all 4 target objects in a specified order in the most efficient way. The meaning of 'efficient' was left to participant to decide: it could mean choosing the shortest path,

or the smallest number of turns or junctions (i.e. navigational decisions). Once the navigational task was finished and all targets were collected, the participants were instructed to stay at the location of the last target and to point to the Start location and then all other target locations twice, in a specified order, using a pointing device. The virtual pointing object in VR looked like a sphere (size of an apple) with an infinitely long ray emanating from it in both directions; the text was displayed next to the hand in VR providing instructions where to point (e.g. 'point to Start'). The targets were placed in such a way that they could not be seen from the pointing location, hence the participant had to rely on their spatial representation of the environment to complete the pointing task. Once the pointing task was complete, the next round of walking and pointing commenced. During all 8 rounds the target locations remained constant.

The 8 rounds were split into 5 rounds of the learning phase and 3 rounds of the test phase. During the learning phase participants always began at the Start location and collected targets in the same sequence Start-Red-Green-Blue-Yellow (S-R-G-B-Y) and finally pointed to all the targets from the yellow sphere location. After that, the walls of the labyrinth disappeared and the participant went directly to the Start location and began the next round.

During the test phase (last 3 round) the navigational sequences were different from the learning phase. Test sequences for scene 1 were as follows: Y-G-B-Y-R, R-B-R-Y-G and G-Y-G-R-B, so that during test the participants had to solve novel navigational tasks. During the test phase participants did not have to go to the start locations in the beginning of a round – they started at the point where the previous round ended. Importantly, the end points in the sequences of the test rounds (i.e. pointing locations) covered all target locations except for the yellow, which was the pointing location during the learning phase. Thus, during test phase we collected pointing data from all target locations, except for the yellow.

Overall there were 7 sessions lasting about 1 hour each. The first session was a practice session during which no data was collected. During this session participants were given instructions and presented with metric VR environments where they could practice walking and pointing in VR. During the other 6 sessions we tested 2 general layouts with 3 repetitions. Between the repetitions (which were carried on different days), the structure of the labyrinth and target locations were identical, but the colours of the targets were changed. This meant that while the instructions remained the same (collect targets in sequence R-G-B-Y) the actual physical paths relating to those tasks were different in different repetitions. In the end, in Scene 1 we had 3 sets of pointing data in the test phase from the location marked as red in Fig. 2 and 2 sets of pointing data from green, blue and yellow locations (in scene 2: 3 sets from blue and 2 sets from the others). Altogether this constitutes 72 pointing vectors per scene per condition per participant: 24 vectors from the red target and 16 vectors from green, blue and yellow.

3 Results

During the first couple of learning trials, while participants were not familiar with the structure of the labyrinth, their paths appeared relatively random, but closer to the end of the learning phase the participants could navigate more systematically and, for

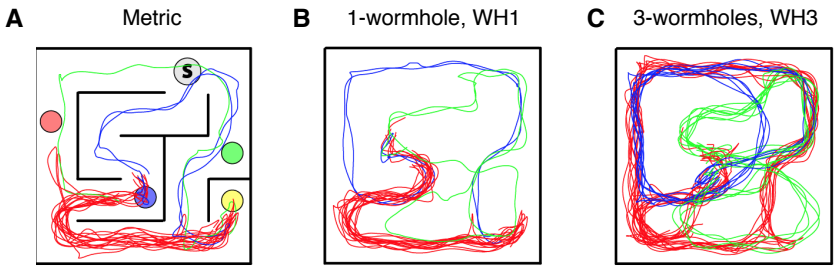

Fig. 3. An example of participants' paths during one subset of the task, "go from B to Y". The full task for this round was: start at Yellow and go to targets in the order G-**B-Y**-R. The shortest path is marked in red, while green and blue lines represent alternative routes. Notice that in the 3-wormhole condition paths are significantly longer.

example, their paths rarely contained loops. Here, we present instead the results of pointing in the test phase, after participants had completed the 5 learning trials. Fig. 3 illustrates the paths that participants took through the maze during the test phase, after the learning phase of 5 rounds. In the cases shown in Fig. 3, participants were always navigating from the blue sphere to the yellow sphere, for a 'metric' (normal) maze, a maze with one wormhole and a maze with three wormholes. There are 3 possible solutions for this task, assuming that participants did not double back on themselves and avoided 'loops' in their trajectory (and, in fact, participants did not do this, demonstrating that they had learned the environment reasonably well by the test phase). The shortest route is marked in red (in these cases, the shortest topological route is also the shortest metric route). Notice that in the 3-wormhole condition the participant paths are significantly longer and more complex than in the metric and one-wormhole conditions even when, topologically speaking, they take the same route as in the metric condition.

Fig. 4 shows data for the pointing task in the labyrinth for Scene 1 (8 participants, 3 conditions and 3 repetitions). We only present data from the test phase, after the 5 learning rounds had been completed. The plot shows signed pointing errors, where positive errors are counter-clockwise relative to the true direction of the target from the centre of the hand-held pointing device. Because of the order in which the pointing zones were visited, the number of 'shots' from each pointing zone was not identical: in scene 1 there were 48 pointing data-points from the red location to each of the other targets, and 32 data-points from green, blue and yellow locations while in scene 2 (see supplementary data online) there were 48 pointing data-points from the blue location and 32 data-points from red, green and yellow locations.

Fig. 4 clearly shows that the magnitude of pointing errors increases along the spectrum from metric to one-wormhole to 3-wormhole condition. This is summarized in Fig. 5a which shows the standard deviation of all the pointing errors (across all participants, repetitions and scenes) for these conditions. However, it would be misleading to describe this as simply an increase in noise or non-specific variability. Fig. 5b illustrates one example of why this is not the case. The solid lines and dotted lines show pointing directions measured on two different days, demonstrating that the

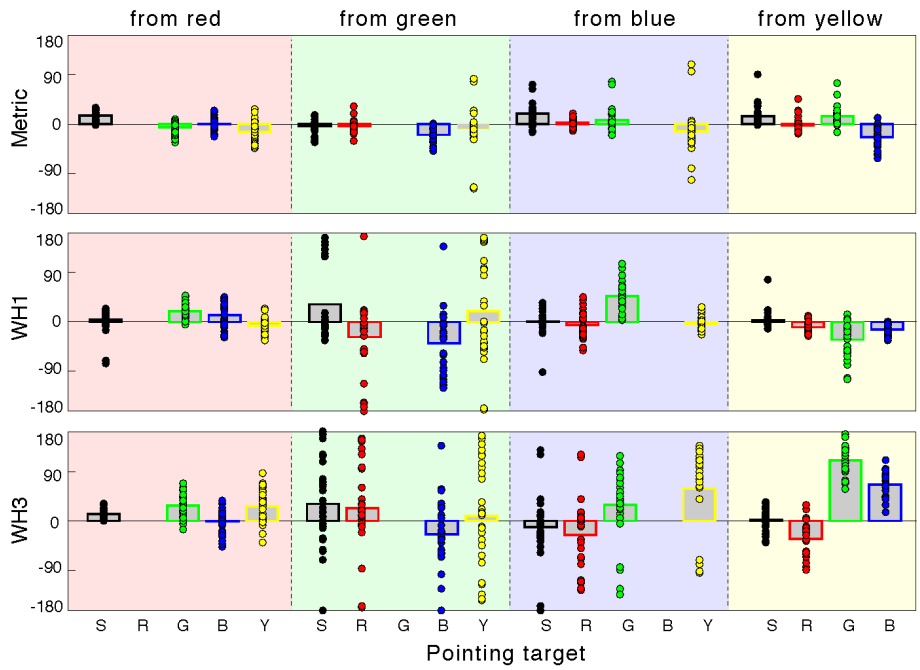

Fig. 4. Pointing errors for all participants (n=8) and all repetitions (3), for Scene 1. Data for the three conditions (metric, one-wormhole and 3-wormhole) are shown in the top, middle and bottom rows. Pointing errors are shown individually with colour symbol indicating the pointing target (Start, Red, Green, Blue and Yellow), the bar height indicating the mean error in each case and the plot background colour indicating the location from which the participant pointed (red, green, blue and yellow columns). Very similar data was obtained for Scene 2.

directions are quite repeatable across days. This makes the errors in the wormhole condition all the more remarkable, since the errors are often close to 180 degrees. In this case, it appears that the participant might have been disoriented, but with the same misestimate of their orientation on both days. However, the pointing vectors cannot be explained by rotation alone – there is a roughly 90° angle between the red and blue pointing vectors while in reality blue and red targets were located on the same line when viewed from green. We discuss some possible ways to model this behaviour in the analysis section.

Such distortions are not limited to a single participant. Fig. 6 shows pointing vectors of all participants to the green target from all pointing locations (left) and the same for the yellow target (right) in the one-wormhole condition. Just like in Fig 6b, the green target was inside a wormhole, while the red, blue and yellow targets and the Start location were not, and hence it is no surprise that the pointing to and from the green target shows particularly marked errors. For example, pointing from the green to the yellow target and from the yellow to the green target both had errors of about 180°. Despite these large errors, participants point accurately to the yellow target from both the red and blue locations which is understandable as neither are in a wormhole and they are connected by 'metric' (physically possible) routes.

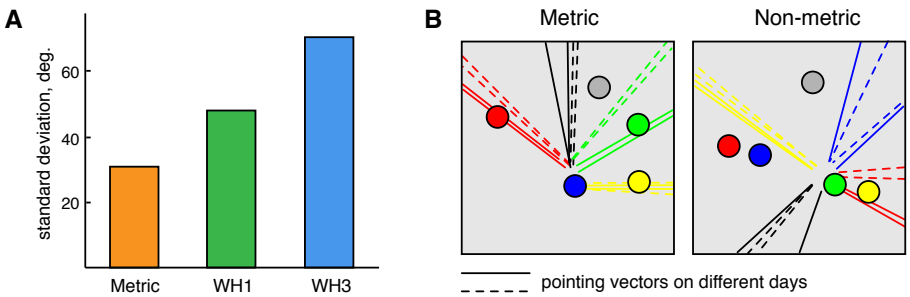

Fig. 5. A) Summary of the pointing errors across all participants and all conditions for both Scenes. B) Examples of pointing vectors in a metric (left) and non-metric condition (1-wormhole). In the non-metric case, the green target was inside the wormhole.

4 Model

The systematic errors that participants make in the pointing experiment indicate that they are not basing their responses on a correct global spatial representation of the target locations. Two qualitatively different kinds of distortions are possible: structural distortion and/or local orientational distortion where the orientation of local reference frames can be misestimated. These are shown in Fig. 7.

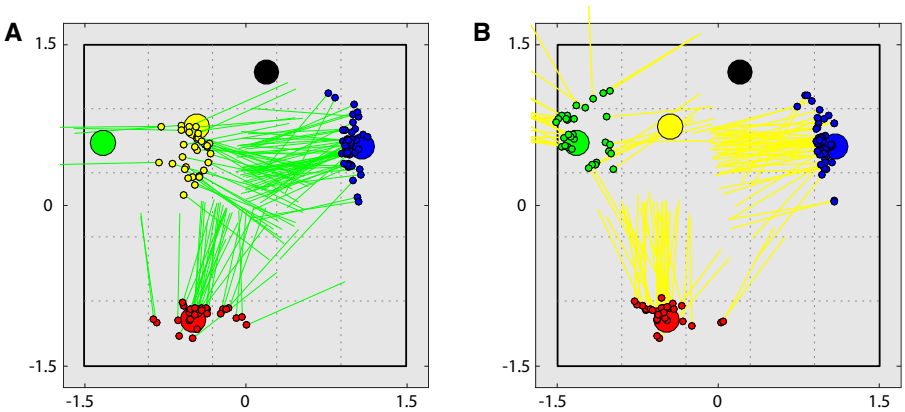

Fig. 6. Example of systematic pointing errors in a non-metric condition with one wormhole. A) Pointing vectors to the green target from each of the other locations. Small circle show the hand location. The green target is inside the wormhole (see schematic in Fig. 10a). Most participants wrongly report that the green target is at the centre of the maze. B) Pointing vectors to the yellow target; these are reasonably accurate from the red and blue locations but there are systematic errors from the green location. These data are from Scene 2.

In order to confirm the presence of distortions and evaluate the relative importance of different kinds of distortions (structural and orientational) we ran the following optimization. We estimated the optimal configuration of targets that best explains the pointing data in the sense that it maximizes the following likelihood function:

$$L(P_R, P_G, P_B, P_Y) = \prod_i^N \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha_i}{\sigma}\right)^2}$$

where P_R, P_G, P_B and P_Y are locations of the targets and α_i is the pointing error with respect to the directions defined by the scene configuration P_R, P_G, P_B and P_Y; the hand location for any given trial was rigidly translated with the target. This function

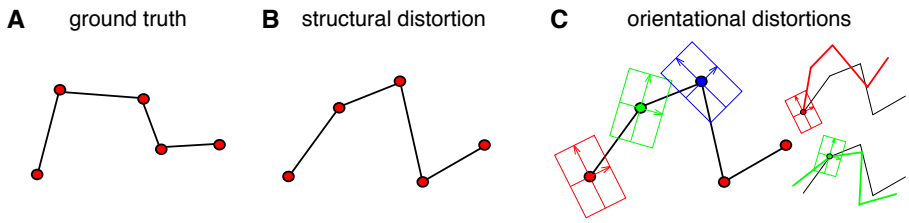

Fig. 7. Distortions of the spatial representations. A) Ground truth structure. B) A structural distortion compared to the ground truth in A). C) Here, in addition, there is orientational distortion so that local reference frames are oriented inconsistently. The insets on the right show the consequences of the local reference frame having a changing orientation.

returns the largest value when angular errors α_i are zero and decreases as values α_i deviate from zero. Altogether there were N = 144 pointing vectors (48 from Blue, and 32 from Red, Green and Yellow) per participant per condition. We used an angular standard deviation for the Gaussian of $\sigma = 50°$ but repeating the analysis with different values showed that the locations returned were rather insensitive to the exact value of σ. The Start location was treated as fixed in the optimization (it was, incidentally, always at a fixed location in the laboratory). We optimised locations of the targets (parameters $P_{R,G,B,Y}$) in order to maximize the likelihood L. The set of potential target locations was discrete and limited to the centers of a 5x5 grid wth the same size as the labyrinth (see Fig. 8). The optimization was done per condition per participant and data from the 3 repetitions were combined together i.e. we assume that on different repetitions participants develop similar representations.

Fig. 8a shows the original pointing vectors from a single participant in Scene 2. The log-likelihood of the pointing data for this configuration according to our model is log(L)=-105 whereas, when the configuration of the targets is optimised to best explain the pointing data, the log-likelihood improves to log(L) = -58 (see Fig 8b). Notice that the green and yellow targets have now (roughly) swapped places, which would help to explain the large errors for pointing between these two targets shown in Fig 6.

Finally, in order to capture the orientational distortions illustrated in Fig 5b, we added an additional component to the model that allowed a rigid rotation of all the parameters, $\rho_r, \rho_g, \rho_b, \rho_y$, describe the orientations of the local reference frames

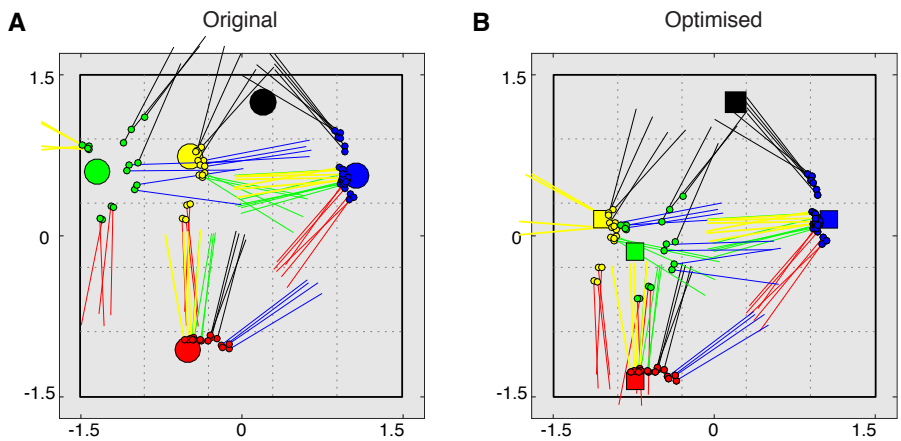

Fig. 8. Example of a distorted spatial map explaining pointing data in a 1-wormhole condition. A) The original configuration of targets and the pointing vectors from 3 repetitions for a single participant. Taking ground truth as the 'model', the log likelihood of these pointing data is log(L) = -105.3 (see text). B) The optimised (distorted) configuration of targets that best explains the pointing data, log(L) = -58.2 (see text). Notice that the location of the green and yellow targets are swapped compared to (A). These data are for Scene 2.

pointing vectors that a participant made from each pointing location. These rotational parameters, $\rho_r, \rho_g, \rho_b, \rho_y$, describe the orientations of the local reference frames around the red, green, blue and yellow pointing locations respectively. Orientations of the local reference frames could only take discrete values of 0°, 90°, 180° and 270°. This was because these were the only directions that a participant could face while looking along a corridor in the labyrinth. The rotational parameters were optimised at the same time as the positional ones. Separate optimisations were carried out per participant (n=8), per scene (n=2) and per condition (metric, one-wormhole and 3-wormhole). Each of the 4 locations was encoded by 2 parameters (x, y) thus, altogether for the first model we had 8*2*4*2= 128 parameters per condition. In the second model, there were 8*2*4*3 = 192 parameters per condition since each location was now encoded by 3 parameters (x,y,ρ) rather than just (x,y). The total likelihood of all the data under each of the two models was calculated by multiplying individual likelihoods per 'shot'. Adding more parameters to a model inevitably improves the likelihood of the best fit of that model, so in order to compare models with different numbers of parameters we calculated the Akaike information criterion (AIC) which penalises the addition of extra parameters as follows:

$$AIC = -2\log(L) + k\,p$$

where L is the total likelihood of all data, p is the number of parameters (0 for original data, 128 for the translation model and 192 for translation and rotational model) and k=2 by definition.

Values for AIC are shown in Fig. 9 for the original data and for these two models for the 'metric' environment, the one-wormhole and the three-wormhole conditions. In

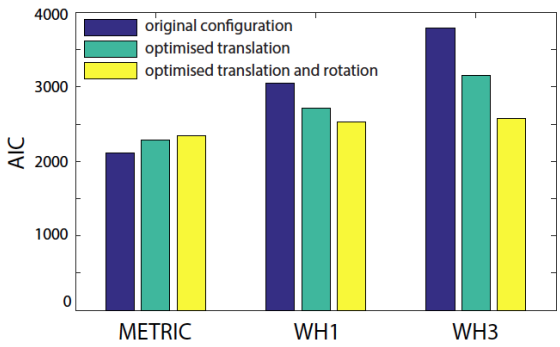

Fig. 9. Akaike Information Criterion values are shown for each condition under 3 models: ground truth, optimised translation and optimised translation-and-rotation).

the 'metric' (normal) maze, allowing the position of the targets and the orientation of the local reference frames to vary does not help: AIC values go up. However, in the one-wormhole condition the pattern is reversed and the trend is even clearer for the 3-wormhole condition. In these cases, the best explanation of the data is one that not only allows for a distortion of the location of the targets in the participant's memory but also assumes that the reference frame they use for pointing may be rotated independently for different pointing locations.

5 Discussion

Our results show that adding wormholes to a labyrinth environment increases pointing errors (Fig. 5a) in a way that is partly explained by assuming that observers build a distorted model of the environment (Fig. 8) but more fully explained by assuming that observers also rotate their local reference frame when they are at certain locations (Fig. 9). The latter explanation is not compatible with a single globally consistent 3D representation of the environment since that could not tolerate different rotations for different parts of the scene.

The argument that 'visual space' is distorted is a familiar one, although much of the literature on this refers to the perception of an observer that is either stationary or only moving a few centimetres [25–28]. It is more unusual to try and find a distorted map of an environment to explain an observer's behaviour when the observer is free to move around and explore all of that environment, partly because in this situation it is not immediately obvious what systematic distortions could apply. Nevertheless, in the peculiar wormhole environments that we have examined, systematic distortions of the scene layout do provide a better explanation of pointing behaviour than the ground truth layout (Fig. 9).

The modelling we have done is agnostic about the causes of the distortions. One speculation is shown in Fig. 10. This illustrates the same maze layout as is shown in Fig. 6 and Fig. 8 where participants made large errors in pointing from the green to the yellow target and vice versa. The green target is located inside a wormhole next to the

left wall and hence is, in reality, to the left of the yellow target. However, as Fig. 8b shows, the best explanation of participants' pointing is that they believe the green target to be near the centre of the maze (i.e. to the right of yellow in these plan views). As Figure 10 shows, this would be the case if participants' conception of the wormhole was that it was 'squashed' into a region between the entrance and exit of the wormhole in the rest of the maze. Participants might do this because they have learnt that the red, yellow and blue spheres are near the outer walls, forcing an interpretation in which the wormhole region is within the central part of the maze.

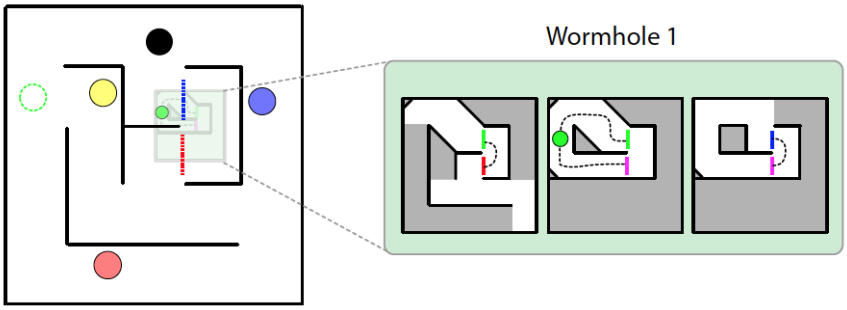

Fig. 10. A speculation about a possible cause of the bias shown in Figs 6 and 8. The corridors making up the wormhole are shown in the green panel on the right. These are shown 'squashed' into the location where the entrance and exit to the wormhole lie (left). This process leaves the green target to the right of the yellow target in the resulting map. This example is taken from Scene 2.

Fig. 10 does not include any distortion of local orientation. When we add this in as a variable (Fig. 9), there is no longer a single, consistent metric representation that corresponds to the output of the model (i.e. x, y and ρ for each target location). This fits with the conclusion of a range of studies based on navigation or spatial knowledge of environments that find evidence against a single consistent representation of 3D space [7–9, 14, 19, 23, 29–31]. For example, Warren et al [5] claimed that there was no consistent representation of space that could account for their data (where the task was to walk directly to targets that had been experienced in a maze with wormholes, i.e. very like our pointing measure but with an indication of perceived distance as well as direction). Our experiment and analysis build on the conclusion of Warren et al (2017) who advocate the idea of a labelled graph. In particular, we make a direct comparison between three models (ground truth, a distorted map or a translation-plus-rotation model) and find that the tanslation-plus-rotation model fits best. This is compatible with the idea that observers learn a topological structure first and then a progressively more accurate labelled graph as they become more familiar with the terrain until, eventually, the information about each edge of the graph is so accurate that the result is, at least in theory, impossible to distinguish from a consistent, metric map. This putative hierarchical calibration process for representing space is similar to hypotheses about observers' representation of surface shape from binocular disparity [32–34].

Acknowledgements. This research was supported by EPSRC/Dstl grant EP/N019423/1.

References

1. Gallistel, C.: The organization of learning. The MIT Press, Cmbridge (1990).
2. O'Keefe, J., Nadel, L.: The Hippocampus as a Cognitive Map. Oxford University Press (1978).
3. Tolman, E.C.: Cognitive maps in rats and men. Psychol. Rev. 55, 189–208 (1948).
4. Meilinger, T., Strickrodt, M., Bülthoff, H.H.: Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation. Cognition. 155, 77–95 (2016).
5. Warren, W.H., Rothman, D.B., Schnapp, B.H., Ericson, J.D.: Wormholes in virtual space: From cognitive maps to cognitive graphs. Cognition. 166, 152–163 (2017).
6. Gillner, S., Mallot, H.A.: Navigation and Acquisition of Spatial Knowledge in a Virtual Maze. J. Cogn. Neurosci. 10, 445–463 (1998).
7. Foo, P., Warren, W.H., Duchon, A., Tarr, M.J.: Do humans integrate routes into a cognitive map? Map- Versus landmark-based navigation of novel shortcuts. J. Exp. Psychol. Learn. Mem. Cogn. 31, 195–215 (2005).
8. Chrastil, E.R., Warren, W.H.: From cognitive maps to cognitive graphs. PLoS One. 9, (2014).
9. Byrne, R.W.: Memory for urban geography. Q. J. Exp. Psychol. 31, 147–154 (1979).
10. Tversky, B.: Distortions in cognitive maps. Geoforum. 23, 131–138 (1992).
11. Chrastil, E.R., Warren, W.H.: Active and passive spatial learning in human navigation: Acquisition of graph knowledge. J. Exp. Psychol. Learn. Mem. Cogn. 41, 1162–1178 (2015).
12. Ishikawa, T., Montello, D.R.: Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. Cogn. Psychol. 52, 93–129 (2006).
13. Meilinger, T., Riecke, B.E., Bülthoff, H.H.: Local and global reference frames for environmental spaces. Q. J. Exp. Psychol. 67, 542–569 (2014).
14. Moar, I., Bower, G.H.: Inconsistency in spatial knowledge. Mem. Cognit. 11, 107–113 (1983).
15. Poucet, B.: Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. Psychol. Rev. 100, 163–182 (1993).
16. Kuipers, B., Byun, Y.-T.: A Robot Exploration and Mapping Strategy Based on a Semantic Hierachy of Spatial Representations. J. Robot. Auton. Syst. 8, 47–63 (1991).
17. Kuipers, B., Tecuci, D.G., Stankiewicz, B.J.: The skeleton in the cognitive map: A computational and empirical exploration. Environ. Behav. 35, 81–106 (2003).
18. Franz, M.O., Schölkopf, B., Mallot, H.A., Bülthoff, H.H.: Learning view graphs for robot navigation. Auton. Robots. 5, 111–125 (1998).
19. Meilinger, T., Newcombe, N.: Spatial Cognition VI. Learning, Reasoning, and Talking about Space. 5248, (2008).
20. Mallot, H.A., Basten, K.: Embodied spatial cognition: Biological and artificial systems. Image Vis. Comput. 27, 1658–1670 (2009).
21. Kluss, T., Marsh, W.E., Zetzsche, C., Schill, K.: Representation of impossible worlds in the cognitive map. Cogn. Process. 16, 271–276 (2015).
22. Vasylevska, K., Kaufmann, H.: Towards efficient spatial compression in self-overlapping virtual environments. 2017 IEEE Symp. 3D User Interfaces, 3DUI 2017 - Proc. 12–21 (2017).
23. Zetzsche, C., Wolter, J., Galbraith, C., Schill, K.: Representation of space: Image-like or

sensorimotor? Spat. Vis. 22, 409–424 (2009).
24. Svarverud, E., Gilson, S., Glennerster, A.: A demonstration of “broken” visual space. PLoS One. 7, (2012).
25. Glennerster, A.: The Time Course of 2-D Shape Discrimination in Random Dot Stereograms. Vision Res. 36, 1955–1968 (1996).
26. Johnston, E.B.: Systematic distortions of shape from stereopsis. Vision Res. 31, 1351–1360 (1991).
27. Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L., Doumen, M.J.A., Todd, J.T.: Exocentric pointing in depth. Vision Res. 48, 716–723 (2008).
28. Ogle, K.: Researches in Binocular Vision. (1950).
29. McNamara, T.P., Diwadkar, V.A.: Symmetry and asymmetry of human spatial memory. Cogn. Psychol. 34, 160–190 (1997).
30. McNamara, T.P.: Mental representations of spatial relations. Cogn. Psychol. 18, 87–121 (1986).
31. Foo, P., Duchon, A., Warren, W.H., Tarr, M.J.: Humans do not switch between path knowledge and landmarks when learning a new environment. Psychol. Res. 71, 240–251 (2007).
32. Tittle, J.S., Todd, J.T., Perotti, V.J., Norman, J.F.: Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. J. Exp. Psychol. Hum. Percept. Perform. 21, 663–678 (1995).
33. Koenderink, J.J., van Doorn, A.J.: Affine structure from motion. J. Opt. Soc. Am. A. 8, 377 (1991).
34. Glennerster, A., Rogers, B.J., Bradshaw, M.F.: Stereoscopic Depth Constancy Depends on the Subject’s Task. Vision Res. 36, 3441–3456 (1996).