# Human pointing errors suggest a flattened, task-dependent representation of space

JENNY VUONG*†, ANDREW W. FITZGIBBON‡, ANDREW GLENNERSTER*†

## Abstract

People are able to keep track of objects as they navigate through space, even when objects are out of sight. This requires some kind of representation of the scene and of the observer's location. We tested the accuracy and reliability of observers' estimates of the visual direction of previously-viewed targets. Participants viewed 4 objects from one location, with binocular vision and small head movements giving information about the 3D locations of the objects. Without any further sight of the targets, participants walked to another location and pointed towards them. All the conditions were tested in an immersive virtual environment and some were also carried out in a real scene. Participants made large, consistent pointing errors that are poorly explained by any single 3D representation. Instead, a flattened representation of space that is dependent on the structure of the environment at the time of pointing provides a good account of participants' errors. This suggests that the mechanisms for updating visual direction of unseen targets are not based on a stable 3D model of the scene, even a distorted one.

## 1. INTRODUCTION

If a moving observer is to keep track of the location of objects they have seen earlier but which are currently out of view, they must store some kind of representation of the scene and update their location and orientation within that representation. There is no consensus on how this might be done in humans. One possibility is that the representation avoids using 3D coordinates and instead relies on a series of stored sensory states connected by actions ('view-based'), as has been proposed in simple animals such as bees or wasps [1, 2, 3], humans [4, 5, 6] and deep neural networks [7, 8]. Alternatively, the representation might be based on 3D coordinate frames that are

---

*Corresponding authors
†School of Psychology and Clinical Language Sciences, University of Reading, United Kingdom
‡Microsoft, Cambridge, United Kingdom

stable in world-, body- or head-centered frames [9, 10], possibly based on 'grid' cells in entorhinal cortex [11, 12] or 'place' cells in the hippocampus [13, 14].

One key difference between these approaches is the extent to which the observer's task is incorporated in the representation. For 3D coordinate-based representations of a scene, task is irrelevant and, by definition, the underlying representation remains constant however it is interrogated. Other representations do not have this constraint. Interestingly, in new approaches to visual scene representation using reinforcement learning and deep neural networks, the task and the environment are inextricably linked in the learned representation [15, 7]. This task-dependency is one of the key determinants we explore in our experiment on spatial updating.

If the brain has access to a 3D model of the scene and the observer's location in the same coordinate frame then, in theory, spatial updating is a straight-forward matter of geometry. It is harder to see how it could be done in a view-based framework. People can imagine what will happen when they move [16, 17, 18] although they often do so with very large errors [19, 20, 21, 22, 23, 24]. In this paper, we examine the accuracy and precision of pointing to targets that were viewed from one location and then not seen again as the observer walked to a new location to point in order to test the hypothesis that a single 3D reconstruction of the scene, built up when the observer was initially inspecting the scene, can explain observers' pointing directions. The task is similar to that described in many experiments on spatial updating such as indirect walking to a target [25, 26, 27, 28], a triangle completion task [29, 30], drawing a map of a studied environment including previously viewed objects' location [31, 19, 32] or viewing a set of objects on a table and then indicating the remembered location of the objects after walking round it or after the table has been rotated [33, 34, 35, 36]. However, none of these studies have compared directly the predictions of a 3D reconstruction model with one that varies according to the location of the observer when they point, as we do here.

Spatial updating has been discussed in relation to both 'egocentric' and 'allocentric' representations of a scene [37, 38, 39, 40] and, in theory, either or both of these representations could be used in order to point at a target. An 'egocentric' model is assumed to encode local orientations and distances of objects relative to the observer [41, 16, 33, 38] while an allocentric model is world-based reflecting the fact that the relative orientation of objects in the representation would not be affected by the observer walking from one location to another [40]. People might use both [37, 17, 39]. Wang and Spelke [38] have emphasised consistency as a useful discriminator between the models. So, for example, disorientating a participant by spinning them on a chair should affect pointing errors to all objects by adding a constant bias if participants use an allocentric representation. The argument that Wang and Spelke [42] make about disorientation conflates two separate issues,

one about the origin and axes of a representation (such and 'allocentric' or 'egocentric') and the other about the internal consistency of a representation. In this paper, we focus on the latter issue. We ask whether a single consistent, but possibly distorted, 3D reconstruction of the scene could explain the way that people point to previously-viewed targets.

## 2. RESULTS

Our stimulus was designed so that participants could get a 'normal' view of the spatial layout of 4 target objects whose position they would have to remember i.e., in either the real world or in a head mounted display, they had binocular view of the scene and could move their head freely (typically, they moved ±25cm). The targets consisted of four different colored boxes that were laid out on one side of the room at about eye height (see Fig. 1a) while, on the other side of the room, there were partitions (referred to from now on as 'walls') that obscured the target objects from view once the participant had left the original viewing zone ('Start zone'), see Fig. 1d and e. For the first experiment, the scene was set up in a real room (Fig. 1a) and replicated in a virtual scene (Fig. 1b). After viewing the scene, participants walked to one of 3 pointing zones where they pointed multiple times to each of the boxes in a specified order (randomized per trial, see Methods for details).

Fig. 1 shows the layout of the boxes in a real scene (Fig. 1a), a virtual scene (Fig. 1b) and in plan view (Fig. 1c, shown here for Experiment 1). The obscuring walls are shown on the left with a participant pointing in the real scene. The measure of pointing error used was the signed angle between the target and the 'straight ahead' indicated by the pointing device subtended at the tracked center of the pointing device, i.e. participants were asked to 'shoot' at the target boxes, yielding the pointing direction illustrated in Fig. 1d. Although not all participants had experience of shooting in video games or similar, the instruction was understood by all and this definition of pointing error gave rise to an unbiased distribution of errors for shots at a visible target (Figure S6) which was not the case for a considered alternative, namely the direction of the pointing device relative to the cyclopean eye (Fig. 1e).

Fig. 2a–2c illustrate pointing directions made by one participant recorded in one trial. The participant pointed to each of the target boxes 8 times in a pseudo-random order (specified by the experimenter) from one of 3 pointing zones (shown in a, b and c respectively). These plots demonstrate a pattern of biases that was typical across participants: they tended to point too far 'north' relative to the true target positions when they were at the north end of the room (see Fig. 1c for the nominal directions in the room) and too far south when they were in the south end
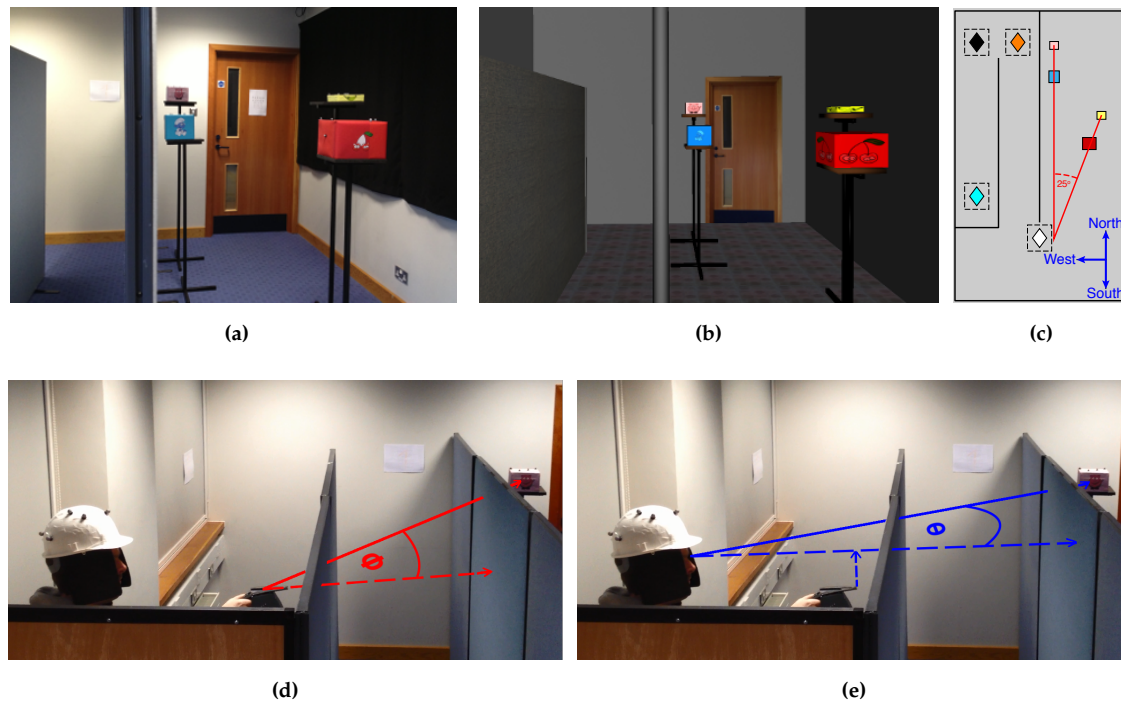
**Figure 1:** *(a) Photo of experimental setup in real world (b) screen shot of experimental setup in virtual reality (c) schematic plan view of one testing layout. 4 boxes are arranged such that the blue and the pink, and the red and the yellow boxes lie along two separate visual directions as seen from the start zone (white diamond). From the start zone, the blue and the red box are always closer than the pink and the yellow box. The two visual directions subtend an angle of 25°. This angle is preserved for all box layouts (though distances varied, see Methods). Pointing to targets was tested at 3 different pointing zones (amber, black and cyan diamond). Black lines indicate positions of walls (in the real world, these were made from temporary walls). The icon in the corner shows nominal 'North' direction. (d) We defined pointing direction as the direction indicated by the pointing device, labeled here as $\phi$. (e) We did* not *use direction subtended at the cyclopean point (midpoint between the left and the right eye along the interocular axis) and the tip of the pointing device, labeled here as $\theta$.*
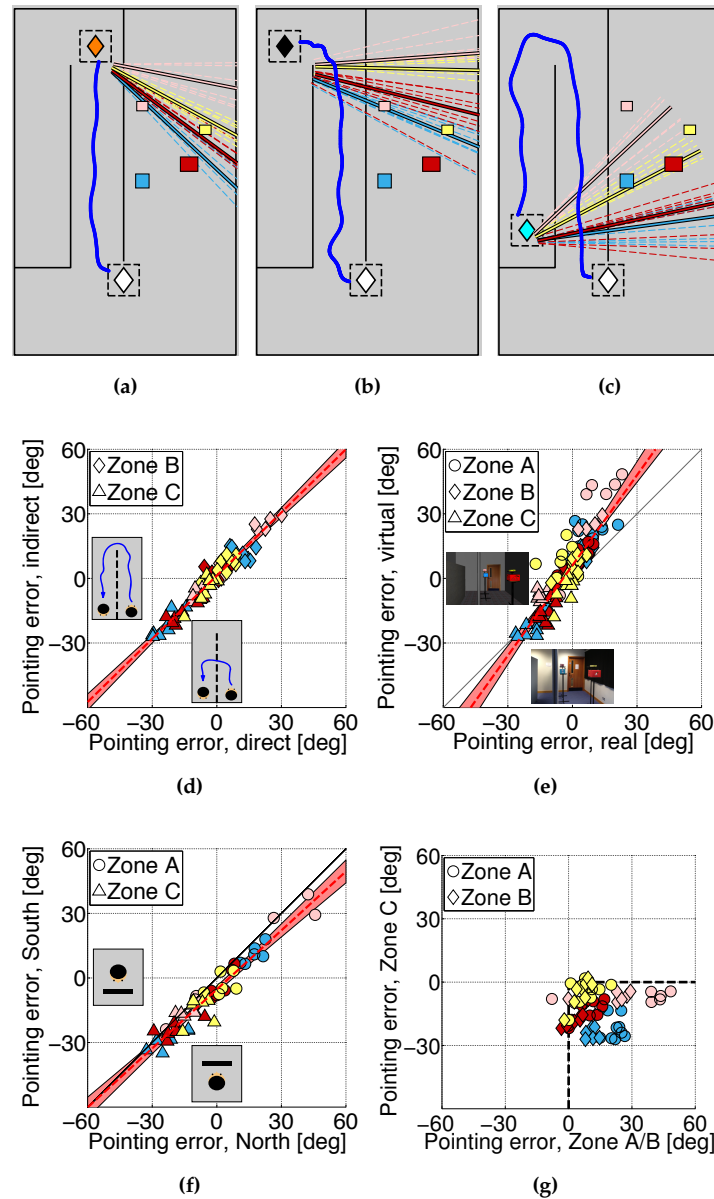
**Figure 2:** *(a-c) Examples of pointing directions of one participant tested in Experiment 1. The white diamond depicts the center of the start zone, (a) the amber diamond shows the center of pointing zone A, (b) the black diamond shows the center of pointing zone B (c) and the cyan diamond depicts the center of pointing zone C. The blue line shows the walking path from the start zone to the pointing zones. (d) Experiment 1: pointing errors for direct and indirect walking paths to the pointing zones compared. Each symbol shows mean pointing error for 20 participants and for a given configuration of target boxes. Colors indicate the boxes, symbol shape indicates pointing zone. (e) Experiment 1: Pointing errors from real world and virtual scene compared. (f) Experiment 2: Pointing errors compared for different initial facing directions. (g) Experiment 1 and 2: pointing errors for Zone C compared to pointing errors for Zone A and B.*

of the room. The magnitude of the errors can be seen in Fig. 2d - Fig. 2g. These plot data for 9 different box layouts, with the colors indicating the color of the box, the symbol type indicating the pointing zone and each symbol showing the mean pointing error for 20 participants. Data for the individual participants are shown in Figs. S2 and S3.

The pointing errors participants made were often large (e.g. $\pm 30°$) and were even larger on the very first trial when (for this trial only) participants were asked to remember the box locations but not told what their task would be (Fig. S5). Pointing directions were highly consistent across conditions as can be seen from the pointing errors for different versions of the experiment shown in Fig. 2d, e and f. Fig. 2d, for example, shows that the pattern of pointing errors was very similar whether participants walked around a set of walls to arrive at the pointing location or walked direct (correlation coefficients 0.92, $p < 0.001$, slope 0.98). There was also a high correlation between pointing errors when participants repeated exactly the same conditions but looked either north or south in order to view the image that told them which was the next pointing target. In these two versions of the experiment, participants' rotation to point at the target was quite different, so the egocentric direction of the target was quite different at the moment they discovered which target they had to point to, but this had no systematic effect on the pointing errors (Fig. 2f, correlation coefficients 0.93, $p < 0.001$, slope 0.91). Finally, there is a high correlation between pointing errors when the experiment was repeated in a real or a virtual environment (Fig. 2e, correlation coefficients 0.88, $p < 0.001$), which supports the findings of Waller and colleagues[43], although the range of pointing errors was greater in the virtual room (slope 1.42). The first-trial data and the data plotted on the ordinate in Fig. 2e were collected in a real environment (Fig. 1a); all the rest of the data were collected in virtual reality. In all these results, the participants' pointing directions were somewhere between the true direction of the target and a direction orthogonal to the obscuring wall (visible in Fig. 1a and shown as a vertical black line in Fig. 2a–2c). This means that pointing errors when participants were in zones A and B (Fig. 2a–2b) were reliably anticlockwise relative to the true target location (positive in Fig. 2g, $M = 12.1, SD = 20.0682, t(949) = 18.7, p < 0.001$ in a two-tailed t-test) while when the participant is in zone C (Fig. 2c) the errors were reliably clockwise (negative in Fig. 2g, $M = -12.9, SD = 14.7, t(479) = -19.2, p < 0.001$).

In order to investigate the origin of these zone-based biases, we set up an experiment in which everything remained constant other than the orientation of the obscuring wall. The scene layout is shown in Fig. 3. As the participant left the start zone, a wall appeared, obscuring the target boxes; it could appear in front of the observer, shown in Fig. 3a and 3d, or slanted as shown in Fig. 3e and 3h. The results are shown in Fig. 3b and 3c. Unlike the manipulations shown in Fig. 2a-2c, the
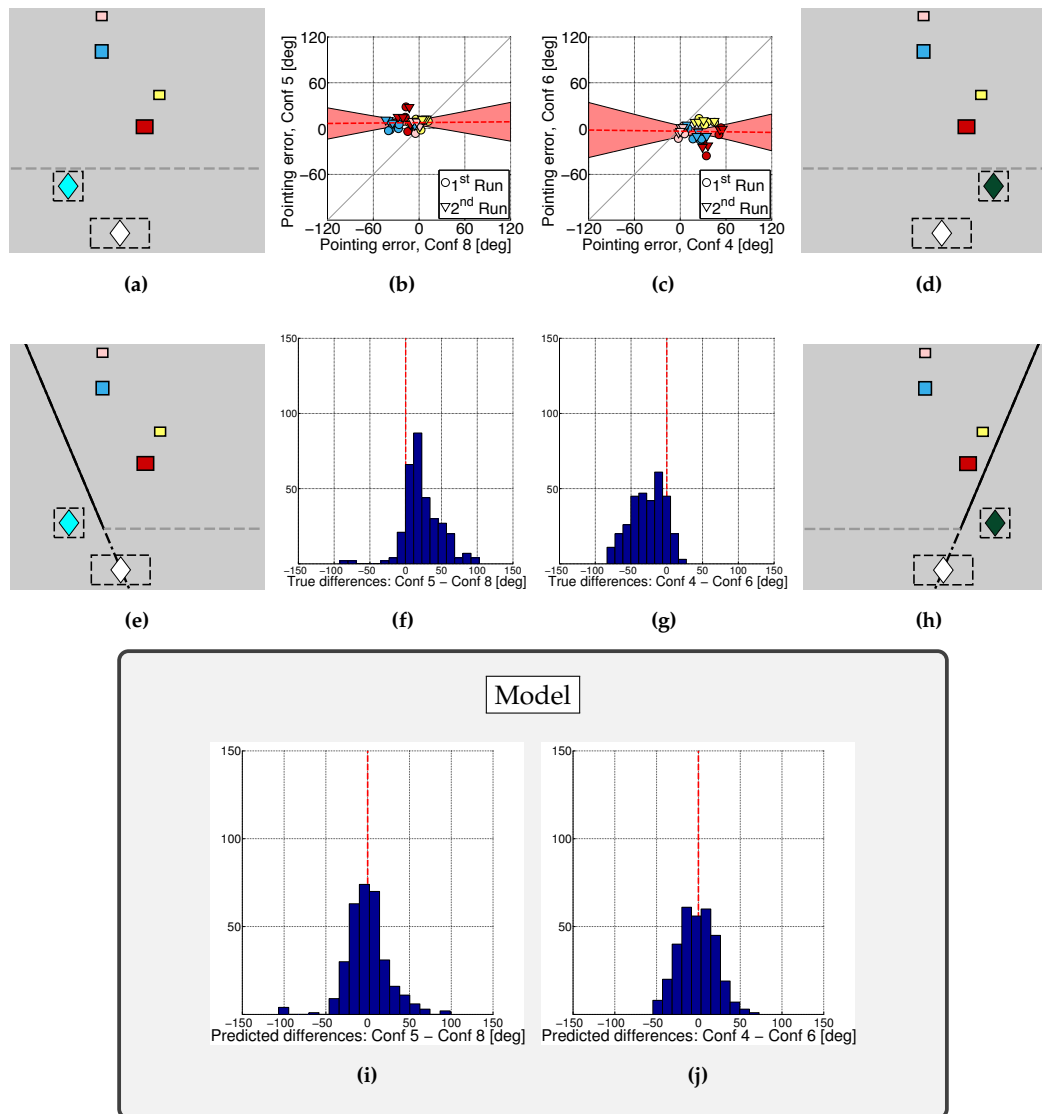
**Figure 3:** *(a,d,e,h) Plan view of a box layout tested in different configurations: The same box positions have been tested at the same pointing zones, but differed in the wall orientation. (b+c) The tight correlations found in Fig. 2 no longer apply here, despite same pointing zones and box layouts. Histogram of actual pointing differences between two wall orientations recorded at (f) zone C and at (g) zone E. Using a one-tailed t-test, the hypotheses could be rejected that the errors come from a distribution centered around 0 (for zone C: $M = 23.5, SD = 26.2, t(319) = 16.0, p < 0.001$; for zone E: $M = -27.1, SD = 23.8, t(319) = -20.3, p < 0.001$). Histogram of predicted pointing directions at (i) zone C (one-tailed t-test, hypothesis could not be rejected ($M = -0.304, SD = 25.4, t(319) = -0.214, p = 0.831$)) and (i) zone E (hypothesis could not be rejected ($M = -1.8156, SD = 21.9614, t(319) = -1.48, p = 0.140$)). For more details about the different configurations, see Fig. S1 in Supplementary Information.*

orientation of the wall had a dramatic effect on participant pointing directions: the data are no longer close to the line of unity as they were in Fig. 2 and have no significant correlation (pointing errors for the scenes shown in Fig. 3a and Fig. 3e are compared in Fig. 3b and have a correlation coefficient of 0.02, $p = 0.921$; equivalently, for the scenes illustrated in Fig. 3d and Fig. 3h, pointing errors are compared in Fig. 3c and have a correlation coefficient of $-0.02$, $p = 0.915$). So, the obscuring wall changes pointing. The way it does so can be seen in Fig. 3f and Fig. 3g which show the distribution of differences between matched conditions (matched by participant, pointing zone, box layout and target box) where only the orientation of the obscuring wall changes. The means of these distributions are both significantly different from zero (one-tailed $t$-test, $p < 0.001$) and shifted in opposite directions. This is what one would expect if participants tended to point somewhere between the true target direction and a direction orthogonal to the obscuring wall, just as they did in Experiment 1 (Fig. 2g).

## 2.1.  Models

The pointing task requires the observer to update an internal representation of their location and orientation. One possible model of our data is that a cumulative error in this updating process is the critical factor explaining the pointing biases. However, as we have already seen, Fig. 2d shows a very high correlation between biases obtained when participants walked directly (i.e. a short distance) or indirectly (a much longer distance) to a pointing zone which would not be expected if pointing biases were caused by cumulative path integration errors. This conclusion is supported by modeling. The model assumes that participants maintain an accurate representation of the target location but mis-estimate their stride length and the rotation of their body on each step. If this noise has zero mean, then the distribution of pointing errors from the final location is also centered on zero (Fig. S11) but if there is a systematic bias in the estimate of stride length or rotation at each step, then systematic biases in predicted location, and hence in pointing direction, can develop. Fig. 4c and Fig. 4d illustrate this effect for two actual trajectories to zone A (in orange) and zone C (cyan) and, as dotted lines, the trajectories resulting from the cumulative effect of a systematic mis-estimate at each step. These trajectories are the ones that best explain the participants' pointing data (see Supplementary Information for details). The errors predicted by the model are larger for an indirect path to than for a direct path to zone C, as one would expect for a path integration model (Fig. S10) but, as we have seen, this is not the case for human observers (Fig. 2d) and a path integration model does not make any predictions about the effect of changing the orientation of the obscuring-wall which is one of the strongest determinants of
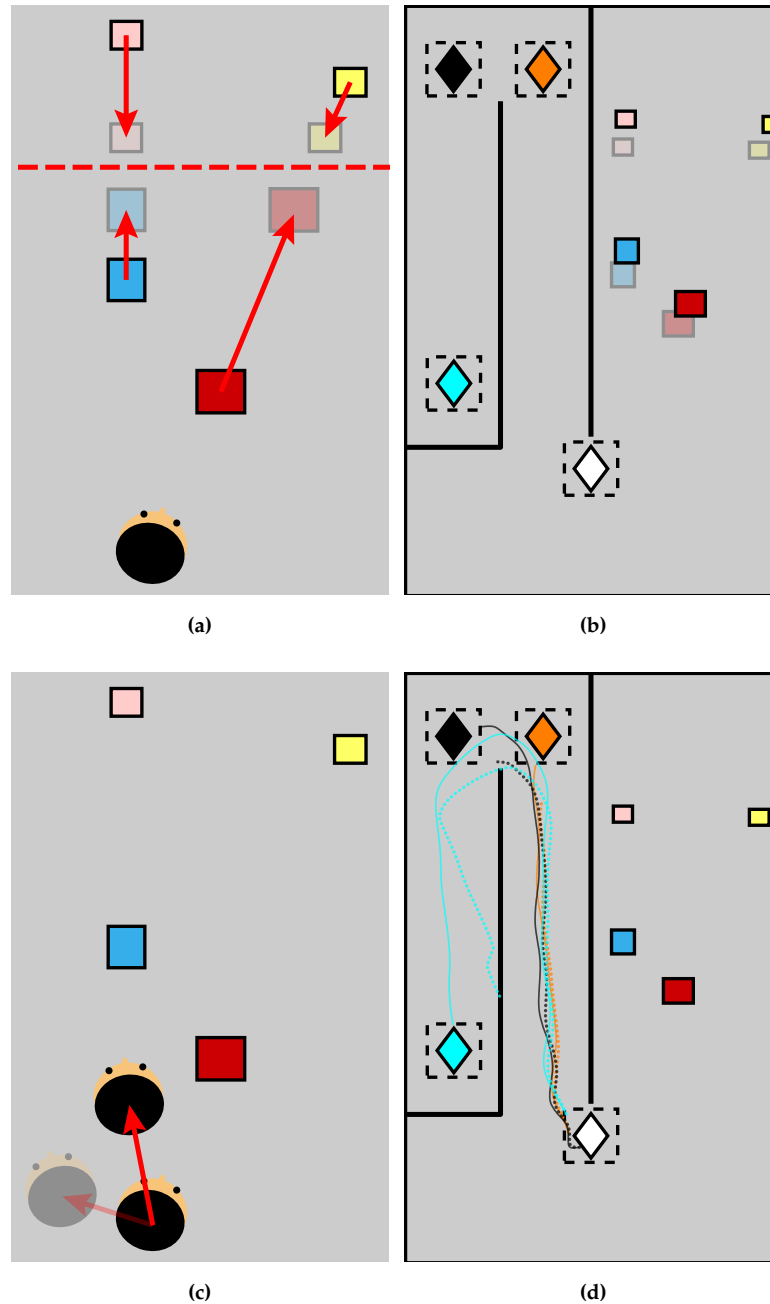
**Figure 4:** *Experiment 1 (a) Schematic explanation of the 'abathic distance' model. (b) Predictions of target box locations under the abathic distance model (opaque rectangles show true box positions). (c) Schematic explanation of the dead-reckoning model. The opaque head shows the true translation; the faded head shows the model translation. The model gives a best fitting translation bias and orientation bias assuming these are the same on every step. (d) Solid lines show the walking paths of a participant, dotted lines show the misestimated walking path using the dead-reckoning model.*
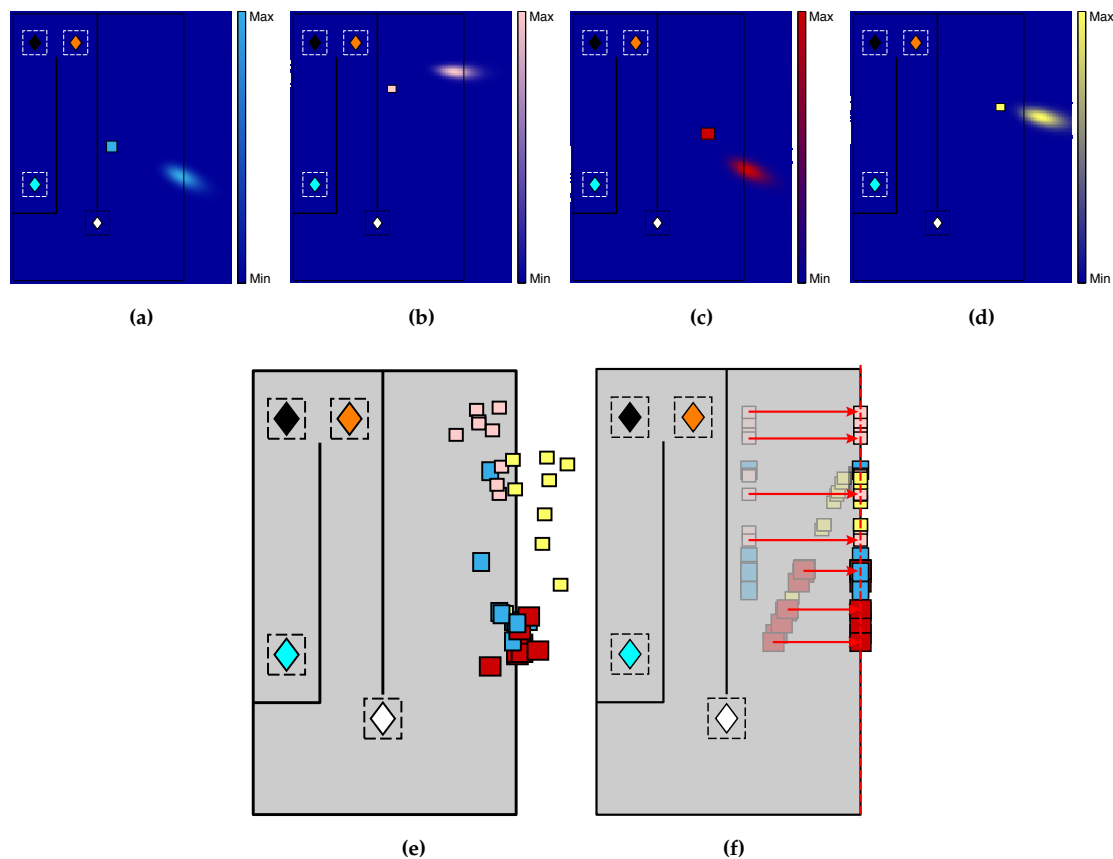
pointing errors in our data (Fig. 3).



**Figure 5:** *(a-d) 'Retrofit' model using the estimated pointing directions of all participants to create a map of where they thought the boxes were, based on their pointing responses. (e) Plan view of the calculated positions of all boxes in each layout using the 'retrofit' model. It seems as if all the boxes are shifted towards a plane parallel to the wall that is obscuring the boxes from where the observer is standing. (f) The projection model assumes that the projection plane is parallel to the wall, with a set distance.*

Another possible cause for systematic errors is that the observer builds a distorted 3D representation of space when they are at the start zone and uses it to guide their pointing. For example, many models of the representation of space built up from binocular disparity assume that there is a distorted mapping between true space and represented space, often with an 'abathic' distance at which objects are judged to be at the correct distance and a compression of visual space towards this plane [44, 45] as illustrated in Fig. 4a. Applying this 2-parameter model to our pointing data, the best-fitting abathic distance is actually about 5m *behind* the observer (see Supplementary Information, Section 4.3.7). and, just like the path integration model, the abathic model cannot

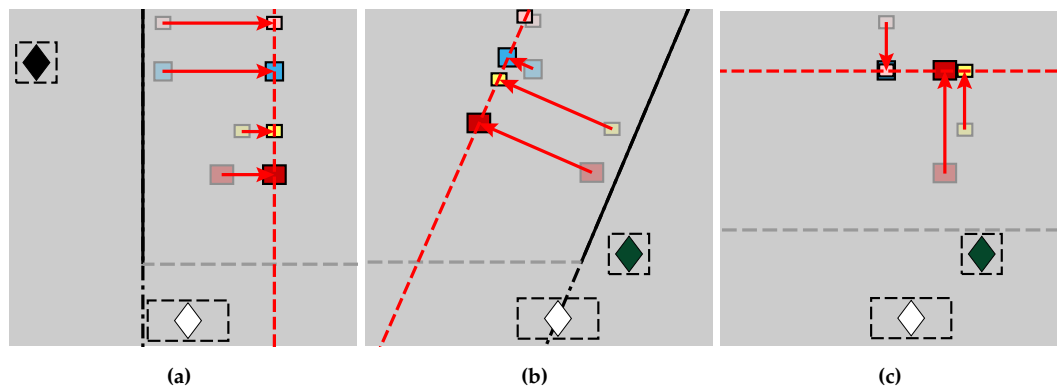account for the effect of the slant of the obscuring-wall.



**Figure 6:** *Schematic explanation of projection model using different wall orientations of Experiment 4*

The abathic distance model is only one case of a model in which the observer uses a distorted 3D model of the scene as the basis of his or her pointing responses. An extreme version of this type of model is to allowing the location of the target boxes to vary freely in order to find a scene layout that maximizes the likelihood of the pointing data, i.e. 'retrofitting' the scene to match the data. The result of doing this for experiment 1 is shown in Fig. 5. We used the pointing directions of all 20 participants from three pointing zones per target box in each condition to calculate the most likely location of the box that would explain their pointing responses. Fig. 5a shows the result of this calculation for the blue box in one condition. It is clear that the derived box location is far from the true location. Figs. 5b, 5c, and 5d show the same for the pink, red and yellow target boxes respectively (see Section 4.3.1 for details and Fig. S8 which shows the same retrofit process applied to the results of experiment 4).

Fig. 5e marks the location of all the peaks of the likelihood distributions illustrated in Fig. 5a-d and all other conditions in experiment 1. These points tend to cluster around a plane that is parallel to the plane of the obscuring wall whereas the true locations of the target boxes (shown in translucent colors in Fig. 5f) are nothing like this. It seems that participants tend to point in a direction that is more orthogonal to the obscuring wall than the true target direction but, in addition, the inferred target locations all have a similar depth relative to the obscuring walls, as if the depth structure of the scene has been 'squashed'. In reality, the pink and blue boxes were very close to the obscuring wall in the center of the room while red and yellow boxes were more distant. Yet, according to participants' pointing data, the *apparent* location of the targets were all at a similar distance relative to the wall. So, as a *post hoc* 'model' of this apparent distortion of the

scene structure, we assume that the observer points to notional targets that are on a plane that is (i) parallel to the obscuring wall and (ii) at a distance determined by the best fit to the points shown in Fig. 5e. For each trial, we project the true box location onto this plane orthographically, as shown in Fig. 5f and then use this notional target location as a predictor of pointing from all zones. In this case, the distance of the projection plane is $1.77m$ from the central obscuring wall. We fix this parameter and use is to determine the distance of the 'model' projection plane in other experiments.

We can now compare the best distorted-world model ('retrofitted' to the data) with this projection-plane model. It might seem counter-intuitive that the 'retrofit' model is not, by definition, the best model in every case since it is fitted to the data but of course, that model is constrained to use the same scene layout independent of pointing zone and wall orientation while the projected plane model is not. Fig. 7 shows the predicted pointing error according to the 'retrofit' model for the data from experiments 1, 2, 3 and 4. Specifically, the most likely layout of the boxes is chosen to explain the pointing data per experiment, i.e. the scene configuration per box layout is fixed per box layout per experiment but allowed to vary between experiments. Fig. 8 shows the same but now using the 'projected plane' model. It is immediately clear that the slopes relating predicted and actual pointing error are steeper in Fig. 8 than in Fig. 7 (correlations 0.86, 0.73 and 0.81 in Fig. 8b, c and d for the 'projected plane' model and 0.54, 0.32 and 0.40 in panels Fig. 7b, c and d for the 'retrofit' model). Note that it is not sensible to compare the models for Experiment 1 (panel Fig. 7a and Fig. 8a) because the 'projected plane' model was generated from the data collected in Experiment 1. As expected given the better correlations, there is a reduction in RMS error due to the projected plane model for Experiment 2 and 4 relative to the retrofit model and, when the number of free parameters in each model is taken into account using Bayesian Information Criterion (BIC) the projected plane model is superior in all three experiments (see Table S2).

Finally, returning to Fig. 3, we can see how the projected plane model explains the bias that is caused by the slanted or East-West wall in that experiment. In Fig. 3j and 3i we have taken paired trials (i.e. with everything else the same except for the wall orientation) and plotted the distribution of pointing 'errors' caused by altering the wall orientation where 'errors' are now defined relative to the *predicted* direction of the target. Unlike Fig. 3g and 3f, where pointing errors were defined relative to the ground truth target direction and the distributions of errors were systematically biased, in both Fig. 3j and 3i zero now indicates the prediction of the projection-plane model and now the mean of the distribution of errors is no longer significantly different from zero.

## 3. DISCUSSION

For a spatial representation to be generated in one location and used in another, there must be some transformation of the representation, or its read-out, that takes account of the observer's movement. The experiments described here show that in humans this process is highly inaccurate and consistently so.

The data appear to rule out several important hypotheses. Crucially, any hypothesis that seeks to explain the errors in terms of a distorted internal model of the scene at the initial encoding stage will fail to capture the strong influence that we have shown of the obscuring wall (see Fig. 3). We examined this hypothesis in relation to the standard models of a compression of visual space around an 'abathic' distance [44, 45] (details in Section 4.3.7 and Fig. S12) but the point is more general. We have allowed *any* type of distortion of the scene that the observer sees from the starting zone to explain the data, provided that the *same* distortion is used to explain a participant's pointing direction from all pointing zones and any wall orientation. The fact that this still provides a worse fit to the data than our projection model (see Fig. 7, 8 and Table S2) is strong evidence against all models of this type.

The data also rule out any of the hypotheses that we have examined based on noisy path integration. One strong piece of evidence in relation to such hypotheses is the high correlation between the magnitude of errors participants make when they walk to a pointing location either via a long or via a short route (Fig. 2d and Fig. S2a). These data suggest that it is the location of the pointing zone, not the route to it, that matters. Also, modeling the data from all 4 experiments, a noisy integration model does a poor job of predicting participant data compared to the other models we tested (Table S1). Instead, our data suggest that people point to a memory of the target boxes that is nothing like where they really are. Our best 'model' of the data is that participants' pointing responses reflect a process that behaves as if the target boxes were further away than they really are and that their apparent locations are compressed towards a plane that is parallel to the wall that obscures the target boxes from the view of the participant (see Fig. 6).
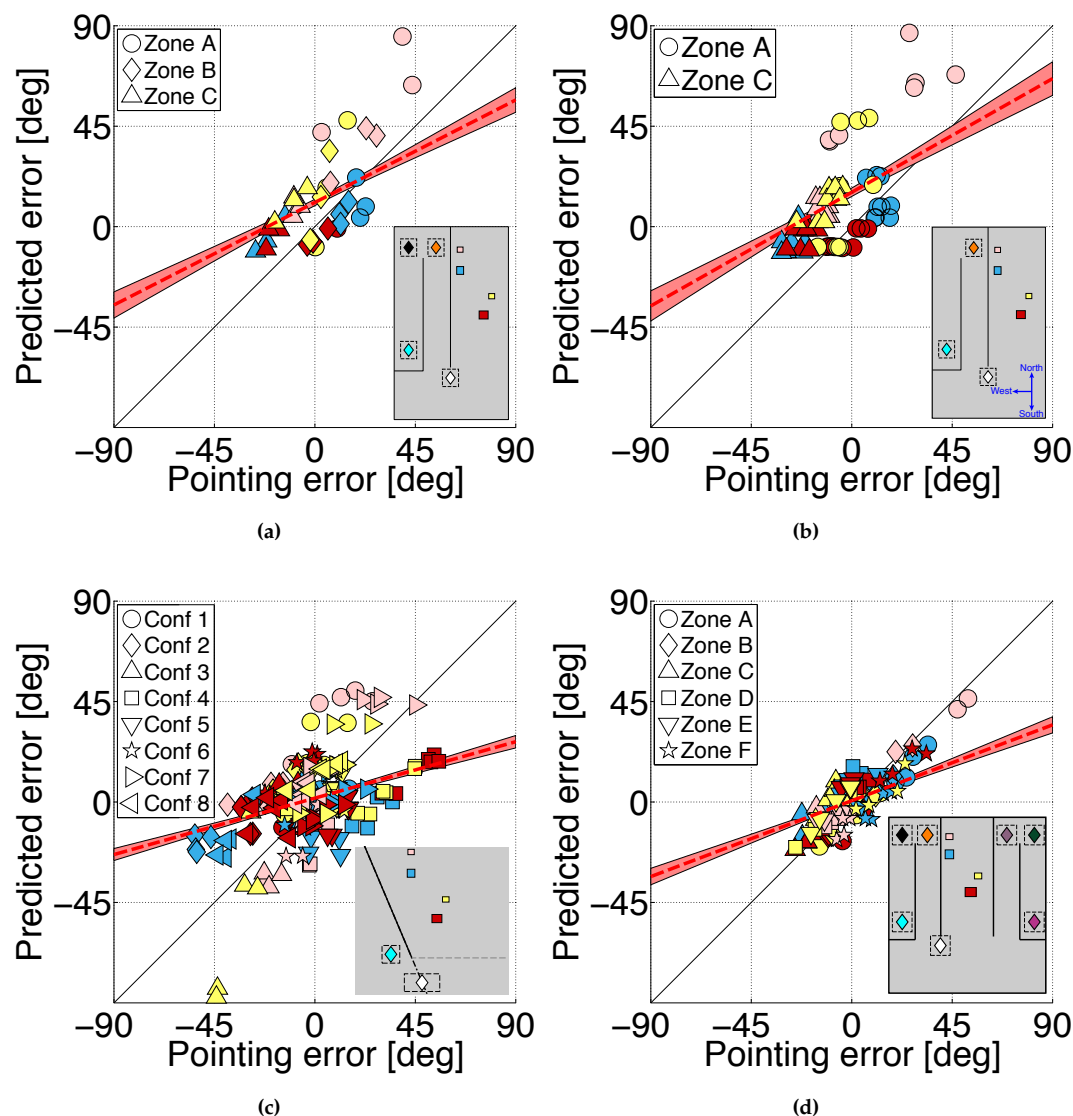
**Figure 7:** *Retrofit Model Experiment 1, 2, 4: Using the retrofit model and the data for Experiment 1, 2, and 4 combined, we can now predict the box positions, see Fig. S7 in the Supplementary material. (a) Experiment 1, R = 0.52 and p < 0.001, RMSE = 24.1°(b) Experiment 2, R = 0.54 and p < 0.001, RMSE = 29.3°(c) for the predicted box positions, R = 0.32 and p = 0.0014, RMSE = 34.06°(d) Experiment 4, R = 0.40 and p < 0.001, RMSE = 25.7°(d) Experiment 3: Retrofit model calculated using all 6 zones to predict box locations, see Fig. S8*
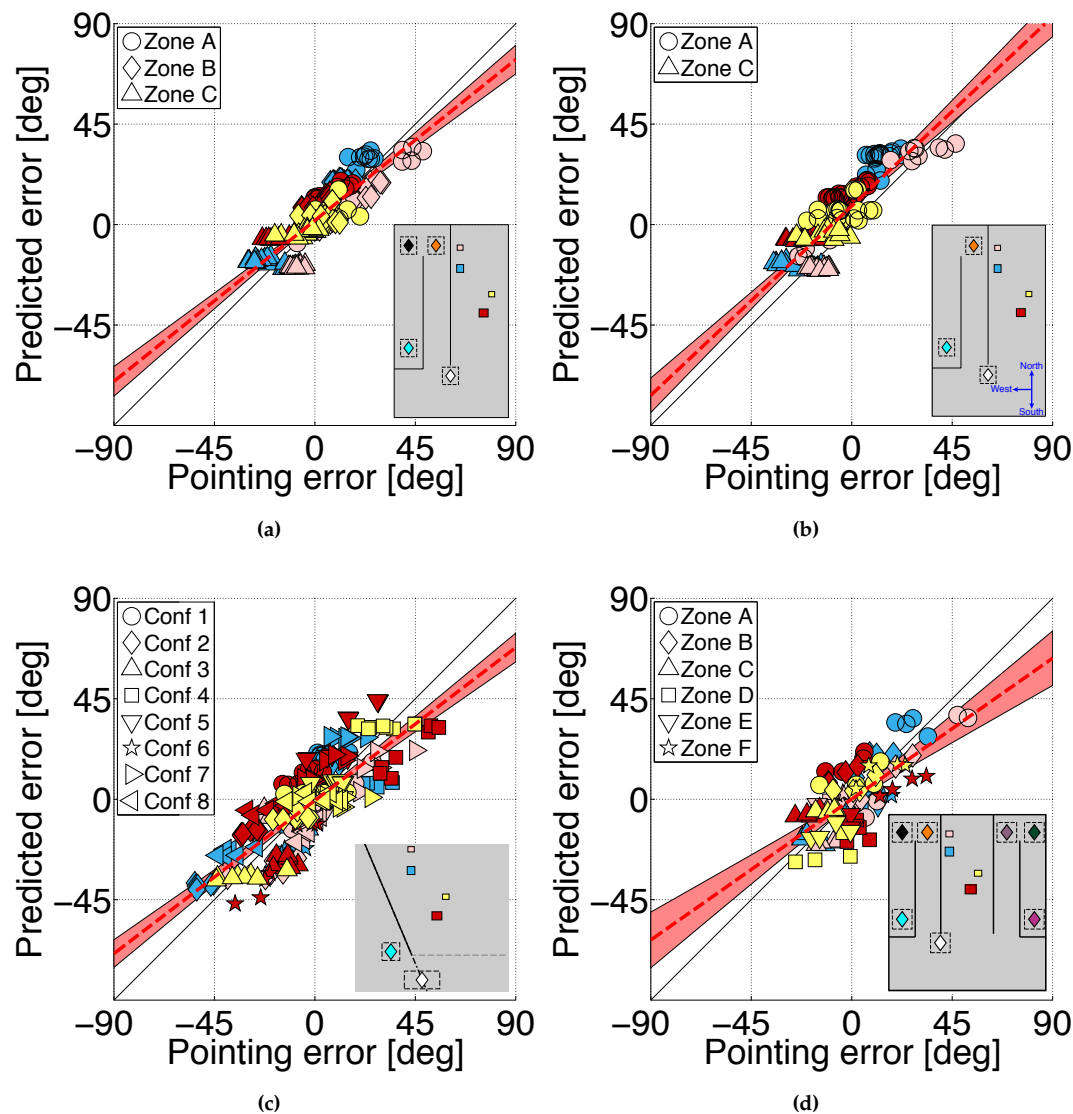
**Figure 8:** *Projection model: (a) Experiment 1, R = 0.86 and p < 0.001 (b) Experiment 2, R = 0.86 and p < 0.001 (c) Experiment 4, R = 0.73 and p < 0.001 (d) Experiment 3, R = 0.81 and p < 0.001*

One paper using a similar paradigm to ours also shows consistent pointing biases [26]. When the pointing zone (in their experiment, called an 'indirect waypoint') was displaced from the location at which the scene layout was learned, pointing was shifted consistently in the direction of the waypoint. The participant was blindfolded during the test phase and guided to the waypoint so there was no equivalent of our wall and only two waypoints were tested so it is hard to make a direct comparison with our data. Nevertheless, their data suggest that the location of the pointing

zone with respect to the start zone has a systematic biasing effect on the direction that participants point to learned targets.

It has often been shown that the observer's orientation can have a significant effect on pointing performance, for example when the orientation of the participant in the test phase differs from their orientation during learning of the scene layout whether the test phase requires real walking [46] or imaginary movement [17, 47]. Meilinger et al. [48] investigated the effect of adding walls to an environment and shown they have a significant effect. However, they do not examine the biasing effects of moving to different pointing locations nor can they be used to deduce any systematic biases in pointing directions since they report 'absolute errors' in pointing, as do many other papers in the field, a measure that conflates variable errors and systematic biases [49, 24, 17, 21, 50, 22, 51, 19]. Röhrich et al. [32] showed that a participant's location (in a town) with respect to target (a well-known market square) was important in determining the orientation at which they drew a plan-view sketch of the square but they did not predict biases in pointing. The fact that we found biases that were large and consistent, whether participants began a trial facing in one direction or the opposite direction (Fig. 2f) is the most relevant evidence from this experiment that egocentric factors did not play a part in causing the biases [17, 39].

The fact that the best 3D reconstruction model does not do well in explaining our data raises the question of what participants do instead. The 'projection plane' model is essentially a re-description of the data in one experiment, rather than an explanation, albeit one that then extends successfully to other situations. Whatever heuristic the visual system uses, it seems to ignore critical aspects of geometry, such as relative depths, so that objects are treated in a more simplistic way (equivalent to assuming that the target objects all lie on a plane) than would be the case if the internal model behaved like a real 3D model. Recent findings using 'Generative Query Networks' [8] suggest that our understanding of the process of 'imagining' novel views of an observed scene, without using a 3D reconstruction, is likely to improve rapidly.

Finally, it is worth considering what effect such large biases might have in ordinary life. The most relevant data from our experiments in relation to this question are, arguably, the pointing biases that we recorded from the very first trial for each participant. We made sure that for this trial that the participant was unaware of the task they were about to be asked to do. Biases in this case were even larger than for the rest of the data (Fig. S5). If these data reflect performance in daily living one might ask why we so rarely encounter catastrophic consequences but, in fact, the task is an unusual one and under most circumstances it is likely that visual landmarks will help to refine direction judgments *en route* to a target.

## Conclusion

Our conclusions are twofold. First, although human observers *can* point to remembered objects, and hence must update some form of internal representation while the objects are out of view, we have shown that they make highly repeatable errors when doing so. Second, the best explanation of our data is not consistent with a single stable 3D model (even a distorted one) of the target locations. This means that whatever the rules are for spatial updating in human observers, they must involve more than the structure of the remembered scene and geometric integration (even with errors) of the path taken by the observer.

## Acknowledgements

## Author contributions

Conceived project: A.G. and A.W.F.; contributed ideas to experiment: A.G., J.V. and A.W.F.; development of testing platform and environments: J.V.; performed experiments and analysis: J.V.; contributed ideas to modeling: A.W.F., A.G. and J.V.; wrote paper: A.G. and J.V.

## Competing interests

The authors declare no competing interests.

## REFERENCES

[1] Cartwright, B. & Collett, T. Landmark Learning in Bees - Experiments and Models. *Journal of Comparative Physiology* **151**, 521–543 (1983). URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Landmark+Learning+in+Bees+Experiments+and+Models#5.

[2] Wehner, R. Desert ant navigation: how miniature brains solve complex tasks. *Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology* **189**, 579–88 (2003). URL http://www.ncbi.nlm.nih.gov/pubmed/12879352.

[3] Collett, M. & Collett, T. S. Local and global navigational coordinate systems in desert ants.

*The Journal of Experimental Biology* **212**, 901–905 (2009). URL http://www.ncbi.nlm.nih.gov/pubmed/19282486.

[4] Gillner, S. & Mallot, H. A. Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience* **10**, 445–463 (1998). URL http://www.ncbi.nlm.nih.gov/pubmed/9712675.

[5] Gootjes-Dreesbach, L., Pickup, L. C., Fitzgibbon, A. W. & Glennerster, A. Comparison of view-based and reconstruction-based models of human navigational strategy. *Journal of Vision* **17**, 11–11 (2017).

[6] Mallot, H. A. & Lancier, S. Place recognition from distant landmarks: human performance and maximum likelihood model. *Biological Cybernetics* 1–13 (2018).

[7] Zhu, Y. *et al.* Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, 3357–3364 (IEEE, 2017).

[8] Eslami, S. A. *et al.* Neural scene representation and rendering. *Science* **360**, 1204–1210 (2018).

[9] Andersen, R. A., Snyder, L. H., Bradley, D. C. & Xing, J. Multi-modal representation of space in the posterior parietal cortex and its use in planning movements **20**, 303–330 (1997).

[10] Snyder, L. H., Grieve, K. L., Brotchie, P. & Andersen, R. A. Separate body- and world-referenced representations of visual space in parietal cortex. *Nature* **394**, 887–891 (1998).

[11] Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–6 (2005). URL http://www.ncbi.nlm.nih.gov/pubmed/15965463.

[12] Killian, N. J., Jutras, M. J. & Buffalo, E. A. A map of visual space in the primate entorhinal cortex. *Nature* **491**, 761 (2012).

[13] O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research* **34**, 171–175 (1971).

[14] Moser, E. I., Kropff, E. & Moser, M.-B. Place cells, grid cells, and the brain's spatial representation system. *Annual review of neuroscience* **31** (2008).

[15] Mirowski, P. *et al.* Learning to navigate in complex environments. *arXiv preprint arXiv:1611.03673* (2016).

[16] Simons, D. J. & Wang, R. F. Perceiving real-world viewpoint changes. *Psychological Science* **9**, 315–320 (1998).

[17] Mou, W., McNamara, T. P., Valiquette, C. M. & Rump, B. Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30**, 142–57 (2004). URL http://www.ncbi.nlm.nih.gov/pubmed/14736303.

[18] Byrne, P., Becker, S. & Burgess, N. Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychological review* **114**, 340 (2007).

[19] Thorndyke, P. W. & Hayes-Roth, B. Differences in spatial knowledge acquired and navigation from maps. *Cognitive Psychology* **589**, 560–589 (1982).

[20] Rieser, J. J. Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**, 1157–1165 (1989). URL http://www.ncbi.nlm.nih.gov/pubmed/2530309.

[21] Presson, C. C. & Montello, D. R. Updating after rotational and translational body movements: Coordinate structure of perspective space. *Perception* **23**, 1447–1455 (1994).

[22] Shelton, A. L. & McNamara, T. P. M. C. Multiple views of spatial memory. *Psychonomic Bulletin & Review* **4**, 102–106 (1997).

[23] Wraga, M., Creem, S. H. & Proffitt, D. R. Updating displays after imagined object and viewer rotations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26**, 151 (2000).

[24] May, M. Imaginal perspective switches in remembered environments: Transformation versus interference accounts. *Cognitive Psychology* **48**, 163–206 (2004). URL http://linkinghub.elsevier.com/retrieve/pii/S0010028503001270.

[25] Rieser, J. J. & Rider, E. A. Young children's spatial orientation with respect to multiple targets when walking without vision. *Developmental Psychology* **27**, 97 (1991).

[26] Klatzky, R. L., Lippa, Y., Loomis, J. M. & Golledge, R. G. Encoding, learning, and spatial updating of multiple object locations specified by 3-D sound, spatial language, and vision. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation cérébrale* **149**, 48–61 (2003). URL http://www.ncbi.nlm.nih.gov/pubmed/12592503.

[27] Medendorp, W. P. Spatial constancy mechanisms in motor control. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **366**, 476–491 (2011).

[28] Bennett, C. R., Loomis, J. M., Klatzky, R. L. & Giudice, N. A. Spatial updating of multiple targets: Comparison of younger and older adults. *Memory & cognition* **45**, 1240–1251 (2017).

[29] Loomis, J. M. *et al.* Nonvisual navigation by blind and sighted: assessment of path integration ability. *Journal of Experimental Psychology: General* **122**, 73 (1993).

[30] 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **31**, 195 (2005).

[31] Kozlowski, L. T. & Bryant, K. J. Sense of direction, spatial orientation, and cognitive maps. *Journal of Experimental Psychology: Human, Perception, and Performance* **3**, 590–598 (1977).

[32] Röhrich, W. G., Hardiess, G. & Mallot, H. A. View-based organization and interplay of spatial working and long-term memories. *PloS one* **9** (2014).

[33] Wang, R. F. & Simons, D. J. Active and passive scene recognition across views. *Cognition* **70**, 191–210 (1999).

[34] Burgess, N., Spiers, H. J. & Paleologou, E. Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences on spatial memory. *Cognition* **94**, 149–166 (2004).

[35] Newell, F. N., Woods, A. T., Mernagh, M. & Bülthoff, H. H. Visual, haptic and crossmodal recognition of scenes. *Experimental Brain Research* **161**, 233–242 (2005).

[36] Mou, W., McNamara, T. P., Rump, B. & Xiao, C. Roles of egocentric and allocentric spatial representations in locomotion and reorientation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32**, 1274 (2006).

[37] Klatzky, R. L. Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In *Spatial cognition*, 1–17 (Springer, 1998).

[38] Wang, R. F. & Spelke, E. S. Updating egocentric representations in human navigation. *Cognition* **77**, 215–250 (2000). URL http://www.ncbi.nlm.nih.gov/pubmed/11018510.

[39] Burgess, N. Spatial memory: How egocentric and allocentric combine **10**, 551–557 (2006).

[40] Meilinger, T. & Vosgerau, G. Putting egocentric and allocentric into perspective. In *International Conference on Spatial Cognition*, 207–221 (Springer, 2010).

[41] Fukusima, S. S., Loomis, J. M. & Da Silva, J. A. Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology: Human Perception and Performance* **23**, 86 (1997).

[42] Wang, R. F. & Spelke, E. S. Human spatial representation: insights from animals **6**, 376–382 (2002).

[43] Waller, D., Beall, A. C. & Loomis, J. M. Using virtual environments to assess directional knowledge. *Journal of Environmental Psychology* **24**, 105–116 (2004). URL http://linkinghub.elsevier.com/retrieve/pii/S0272494403000513.

[44] Johnston, E. B. Systematic distortions of shape from steoreposis. *Vision Research* **31**, 1351–1360 (1991).

[45] Foley, J. M. Binocular distance perception **87**, 411–433 (1980).

[46] Meilinger, T., Riecke, B. E. & Bülthoff, H. H. Local and global reference frames for environmental spaces. *The Quarterly Journal of Experimental Psychology* 1–28 (2013).

[47] Waller, D., Montello, D. R., Richardson, A. E. & Hegarty, M. Orientation specificity and spatial updating of memories for layouts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **28**, 1051–1063 (2002). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.28.6.1051.

[48] 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).

[49] Farrell, M. J. & Robertson, I. H. Mental rotation and automatic updating of body-centered spatial relationships. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **24**, 227–233 (1998). URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0278-7393.24.1.227.

[50] Wang, R. F. Between reality and imagination: When is spatial updating automatic? *Perception & Psychophysics* **66**, 68–76 (2004).

[51] Shelton, A. L. & McNamara, T. P. Systems of spatial reference in human memory. *Cognitive Psychology* **43**, 274–310 (2001). URL http://www.ncbi.nlm.nih.gov/pubmed/11741344.

## 4. SUPPLEMENTARY MATERIAL

## 4.1. General Methods

In virtual reality, participants viewed the stimuli through a binocular head-mounted display NVIS SX111 with horizontal view of $102°$, vertical view of $64°$ and binocular overlap of $50°$. Each display for the right and the left eye had a resolution of $1280×1024$ pixels refreshed at a rate of 60 Hz and an end-to-end latency of 2 frames. The tracking system consisted of 14 MX3/T20/T20-S Vicon tracking cameras, a tracking computer (Quad Core Intel Xeon 3.6GHz processor, NVidia Quadro K2000 graphics card, 8GB RAM) running Windows and a graphics computer (8 core AMD Opteron 6212 processor, dual NVidia GeForce GTX 590 graphics cards, 16GB RAM) running Linux that generated the experimental stimulus. The head-mounted display was tracked at a rate of 240 Hz.

In the real environment, participants wore a tracked hardhat with blinkers on the side to imitate the horizontal field of vision of the head-mounted display. The participants were blindfolded when guided to the start zone by the experimenter. The target boxes had markers stuck to them so that they could be tracked and their position recorded during trials. In both virtual and real environment, participants used a hand-held, fully-tracked pointer resembling a gun, to indicate a pointing direction with the direction of the pointer and with a button press.

All participants had normal or corrected-to-normal vision and could distinguish the colors of the target boxes. Their acuity and stereoacuity were tested before the start of the experiment: all had 6/6 vision or better and normal stereopsis with 60 arc seconds or better.

### 4.1.1 Ethical Approval

The experiments described in this study received the approval of the University of Reading Research Ethics Committee. Informed consent was obtained from all participants.

### 4.1.2 Experiment 1

The first pointing experiment was conducted in both a real-world scene and in virtual reality. Participants were first tested in a real-world block, followed by a block in virtual reality and another real-world block at the end.

**Participants** We recruited 22 participants (aged 19–46) who were paid £10.00 per hour. Data from 2 participants had to be discarded as one failed to understand the task and one felt too

uncomfortable wearing the head-mounted display. 19 out of the 20 were naïve to the experiment, and 1 was a researcher in our lab who had prior knowledge about the task.

**Procedure**    In virtual reality, in order to start a trial participants walked towards a large green cube with an arrow pointing towards it in an otherwise black space then, as soon as they stepped inside the cube, the stimulus appeared. In the real world, participants wore a blindfold and were guided to the start zone by the experimenter. As soon as they were in the center of the start zone, the experimenter removed the blindfold. In both real and virtual experiments, the participants then viewed 4 boxes from this start zone. They were told that from where they were standing, the blue box was always closer than the pink box and both were in the same visual direction. Likewise, the red box was always closer than the yellow box and also both in the same visual direction as viewed from the start zone, see the example plan view in Fig. 1a–c. Participants took as much time as they needed to view and memorize the box positions, although this was typically between 10 and 20 seconds. If they attempted to leave the start zone to walk closer to the target boxes, the whole scene disappeared in virtual reality or were inhibited by the experimenter in the real world. When they had finished memorizing the box layout, they walked behind a 'wall' (partition) towards a pointing zone. In virtual reality, the pointing zone was indicated by a colored poster (colored according to the color of the box to which they should point) and this poster only appeared after they left the start zone similarly, in the real-world task, there were three white posters at the three pointing zones (see Fig. 1c) and the participant was only told which one to stop at after they had left the start zone. In virtual reality, the poster changed its color after each shot to indicate the next target box to point at. In the real world, the experimenter told the participant which box to point at. Following these instructions, the participant had to point 8 times to each of the 4 boxes in a pseudo-random order, i.e. 32 pointing directions in all. The end of each trial was indicated by the whole scene disappearing in virtual reality (replaced by the large green box) and in the real world the experimenter told the participant to wait to be guided back blindfolded to the start. The participant received a score ranging from 0–100% reflecting the accuracy of all the shots but this information could not be used to infer the direction or magnitude of their pointing error on any given shot.

**Box layouts and stimulus**    8 participants were tested on 9 box layouts (27 trials in total – 9 layouts × 3 pointing zones), whereas the remaining 12 participants were tested on 4 layouts (12 trials in total – 4 layouts × 3 pointing zones). All layouts are shown in online Supplementary Data. The virtual and the real stimulus were designed to be as similar as possible, with a similar scale,

texture mapping taken from photographs in the laboratory and target boxes using the same colors and icons, see Fig. 1a–b.

The box positions in the 9 box layouts were chosen such that the following criteria were satisfied, see Fig. 1(c):

- The blue and the pink boxes were positioned along one line (as seen in plan view) while the red and the yellow box were positioned along a second line. The two lines intersected inside the start zone.

- All box layouts preserved the box order: this meant that the blue box was always in front of the pink box, and the red box was in front of the yellow box, but the distances to each box varied from layout to layout.

- The 2 lines were 25° apart from one another. This number allowed a range of target distances along these line while the boxes all remained within the real room.

### 4.1.3 Experiment 2

This experiment aimed to identify the influence of the facing orientation at the pointing zones. 7 participants, all of whom had been participants in the first experiment, viewed the same 6 layouts as in Experiment 1 but in virtual reality only. They then walked to either zone A or zone C, stopped and faced a poster located either to the 'North', 'South', or 'West' of the pointing zone (unlike Experiment 1 when there was only one poster position per pointing zone, see Fig. S1a. Otherwise, the procedure was identical to Experiment 1, with the exception that the participant pointed 6 times (rather than 8) to each of the boxes in a random order at the pointing zones. Data are shown in Figure S3 and Figure 2f.

### 4.1.4 Experiment 3

The third experiment was also carried out only in virtual reality and here the testing room was scaled (in the x-y plane) by a factor of 1.5, i.e. the height of the room did not change. The pointing zones A, B and C were mirrored along the center line of the room to create an additional 3 pointing zones D, E and F, see Fig. S1b but the physical size of the laboratory dictated that participants could only go to one side of the scene or the other (either zones A, B, C or zones D, E, F). The participants did not know whether they would walk to the left or to the right from the start zone until the moment that they pressed the button (when the boxes disappeared). There were 6 participants, 3 of whom had taken part in the previous experiments while 3 were naïve.

### 4.1.5 Experiment 4

The scene layout for this experiment is shown in Fig. S1c–j. There were 4 pointing zones B, C, D and E and only 4 box layouts. All the data was collected in virtual reality. For every condition (box layout and pointing zone), participants carried out 2 trials (12 shots per target) with the wall in one orientation and the same number of trials with the wall in a different orientation.
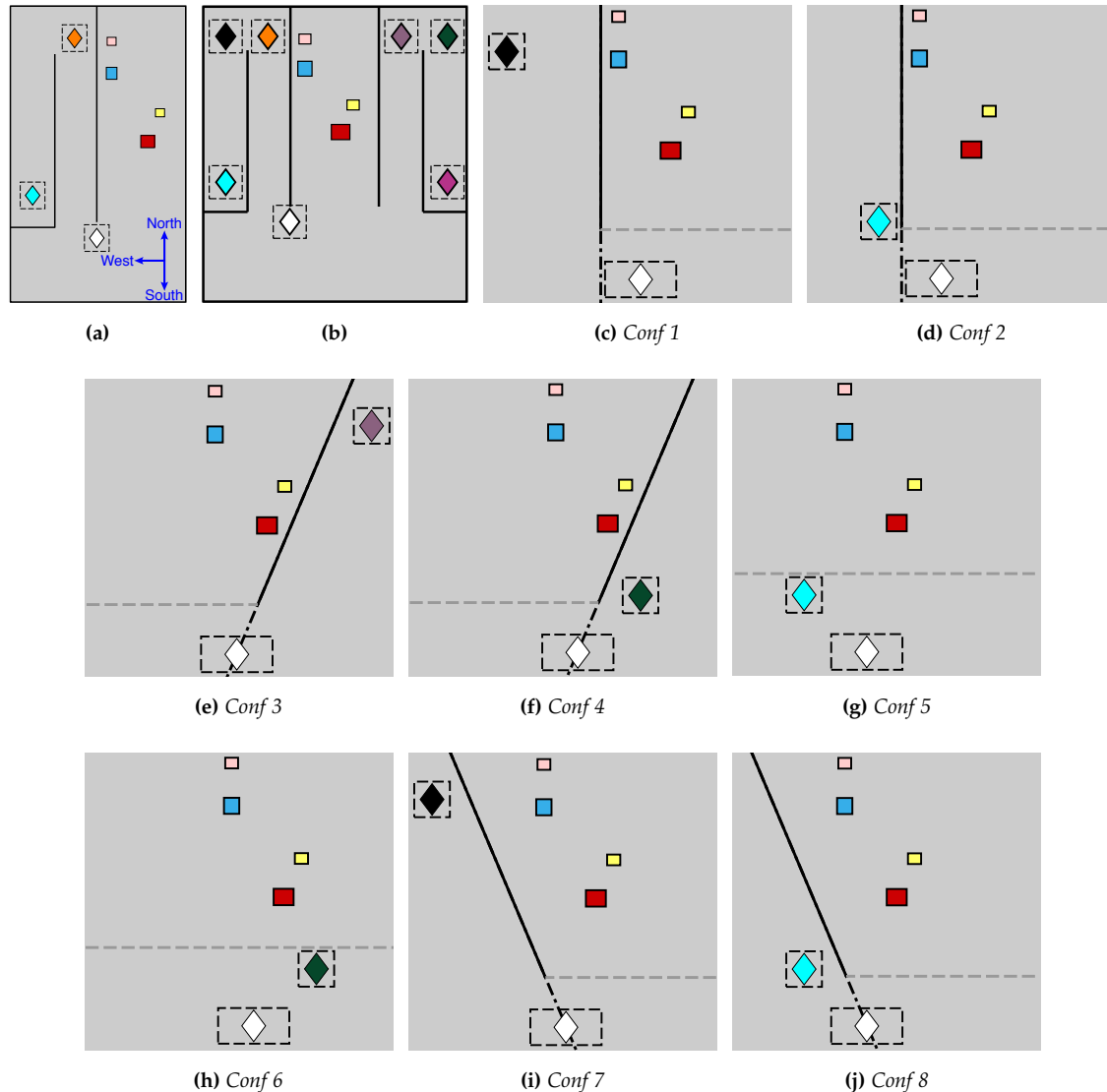
**Figure S1:** *Plan view of (a) Experiment 2, (b), Experiment 3 and (c–j) Experiment 4 with 8 different configurations. Black solid line indicates walls that were always present. In (c–j), the gray dotted line shows the wall that appeared when the participant pressed a button at the start zone (white diamond). The black dotted line shows the wall that appeared after participant left the start zone. In (g) and (h) the wall remained in the same place.*

4 naïve and 6 non-naïve were recruited for this study. The procedure of this experiment was similar to Experiments 1–3 with the following differences: Participants were standing at the start zone and viewed the target box positions. After pressing a button, a 1.8m tall wall appeared in front of them, see gray dotted line in Fig. S1c–j. A white poster at eye height and a green square

on the floor appeared in the room indicating the location of the pointing zone and the orientation. When the participant left the start zone, another wall appeared, see black solid line in Fig. S1c–j. The rest of the trial was as described in Experiment 1–3.
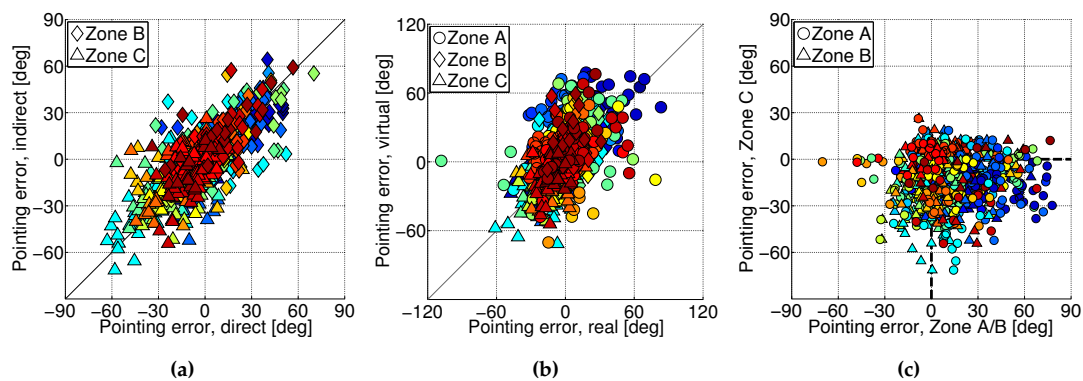
## 4.2. Results

### 4.2.1 Individual participant data



**Figure S2:** *Experiment 1 — Like Figure 2 but showing data for individual participants. (a) Pointing error collected in a real and virtual environments; (b) comparison of pointing errors after direct and indirect walking. Symbol colors indicate different participants.*
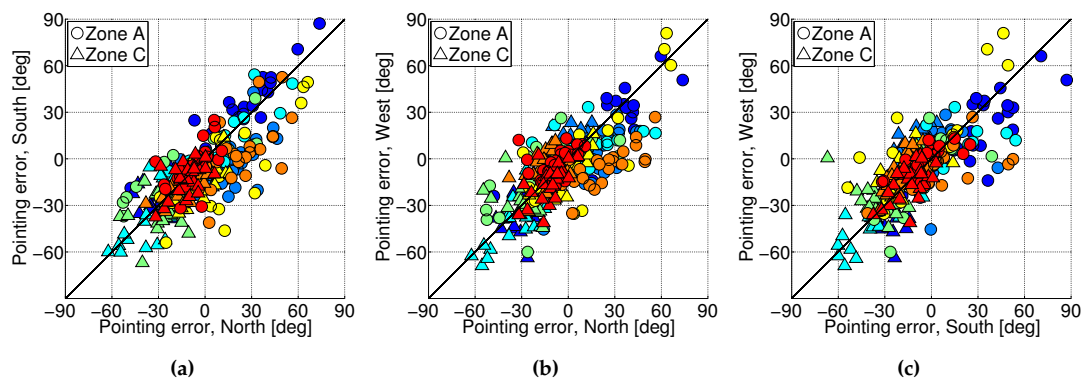


**Figure S3:** *Experiment 2 — Pointing errors collected with different facing orientation, color-coded by participants. (a) Participants' errors when their initial orientation was 'North' plotted against errors for 'South' orientation, (b) 'North' compared to 'West' and (c) 'West' compared to 'South'. See Figure 1c for icon showing nominal 'North' in the scene.*
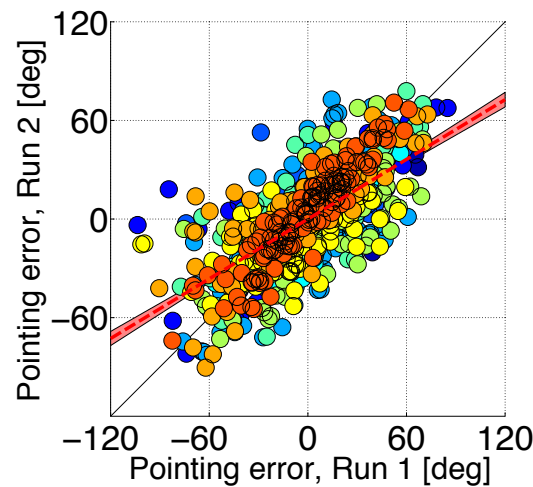
**Figure S4:** *Experiment 4 — test/re-test reliability. In this experiment, the same conditions were repeated. Errors on the second run are plotted against errors on the first run. The slope of the correlation is 0.61, indicating that participants had a smaller range of pointing errors in the second run of the experiment. As in previous figures, each color indicates a different participant.*
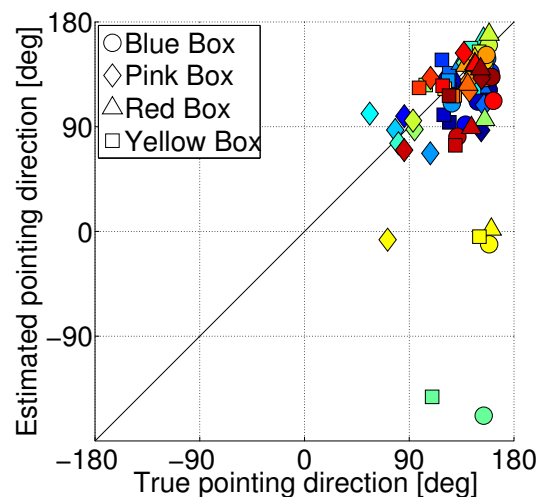
### 4.2.2 First Trial Data



**Figure S5:** *Data from participants' very first trial (in Experiment 1). Participants viewed the stimuli in the real world without further instructions. After walking to the pointing zone, they were asked to point to the four boxes in a random order, eight times each (i.e. 32 shots per participant). Colors indicate different participants as in Fig. S2*

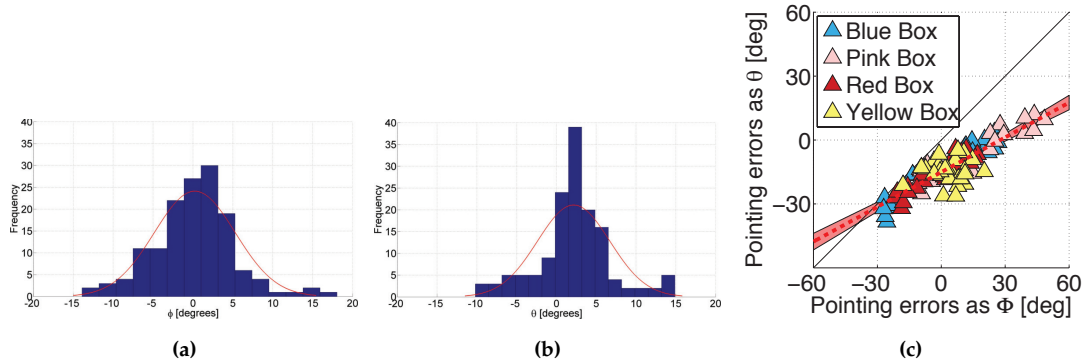### 4.2.3 Alternative definitions of pointing direction



(a)  (b)  (c)

**Figure S6:** *Evidence supporting the choice of 'shooting' angle ($\phi$) rather than visual direction ($\theta$) as the most appropriate definition of pointing direction. In a control condition, participants pointed at target boxes from the start zone so the target was visible (18 participants tested on 2 box layouts with 4 target boxes in each layout, 144 shots in total). Their instruction was to 'shoot at the box', just as it was in the spatial updating experiments. (a) Distribution of pointing errors when target direction is defined relative to the pointing device and shooting direction is defined by the orientation of the device (see $\phi$ in Fig. 1d). (b) As for (a) but with pointing errors defined as the difference in visual direction of the target and the pointing device as measured from the cyclopean point (see $\theta$ in Fig. 1e). The mean of the distribution is significantly biased for $\theta$ (t-test, $p < 0.001$) but not for $\phi$ ($p = 0.605$) suggesting $\phi$ reflects participants' intentions when pointing. (c) There is a significant correlation between the $\phi$ and $\theta$ measures in the spatial updating experiments (R = 0.91). These data are from Experiment 1. When $\phi = 0$, $\theta$ is about $-20°$. This is the direction of bias that would be created if a right-handed participant held the pointing device slightly out to their right and pointed to a target directly ahead of them.*

## 4.3. Visual Models

### 4.3.1 Retrofit model

With $\mu = 0$ and an arbitrarily chosen $\sigma = 15$, it is now possible to calculate for each possible box position in $B$ the likelihood of the participant pointing to that possible box position using the probability density function:

$$L_{p,m}^{b,l,k} = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(a_k - \mu)^2}{2 \cdot \sigma^2}} \tag{1}$$

with $p = [1, \ldots, 20]$ for 20 participants, and $b = [1, \ldots, 4]$ for all 4 boxes in $l = [1, \ldots, 9]$ 9 box layouts, and $m = [1, \ldots, 3]$ for all 3 testing zones, and $k = [1, \ldots, 4\,000]$ for all possible box

positions defined in $B$, The maximum likelihood is then calculated for each box in each layout across all participants across all zones:

$$L^{b,l,k} = L^{b,l,k}_{1,1} \cdot L^{b,l,k}_{1,2} \cdot L^{b,l,k}_{1,3} \cdot L^{b,l,k}_{2,1} \cdot \ldots \cdot L^{b,l,k}_{P,M} \tag{2}$$

with $P = 20$ for the total number of participants and $M = 3$ for the total number of shooting zones.

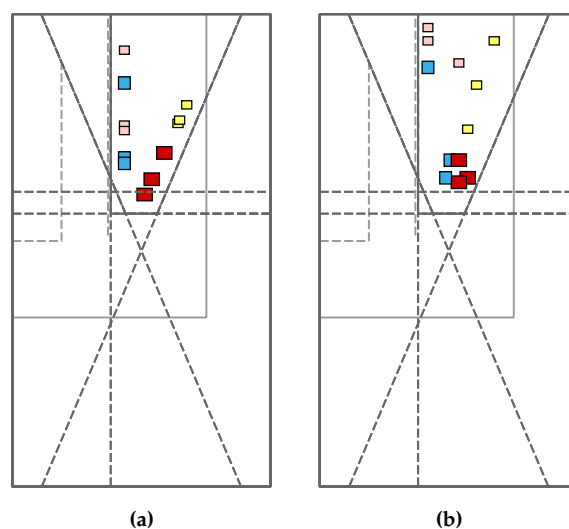### 4.3.2 Retrofit target locations for Experiment 1, 2 and 4 combined



(a)                           (b)

**Figure S7:** *Box layouts tested in Experiment 1, 2, and 4: (a) Real box positions of 3 box layouts plotted in the same plan view. (b) Predicted box layouts calculated using the retrofit model, based on combined data from all studies.*

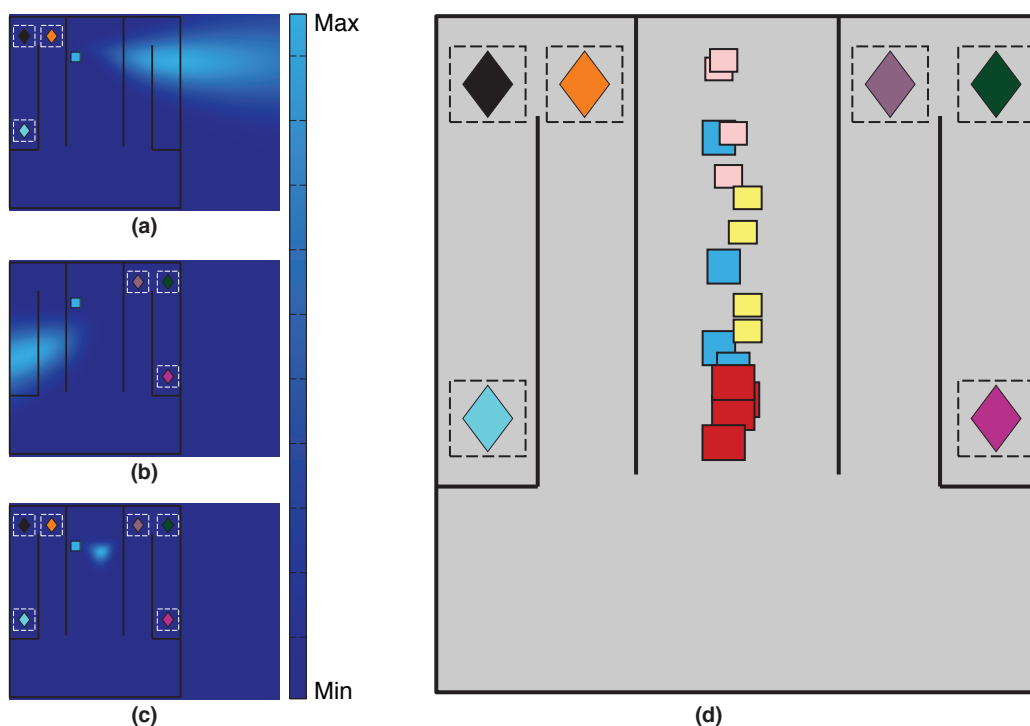### 4.3.3 Retrofit Model Experiment 3



**Figure S8:** *Experiment 3, 'retrofit' model using the estimated directions of one participant pointing to the blue box in one layout (a) at zones A–C alone or (b) at zones D–F alone or (c) A–F all together. (d) Similar to (c) but now showing the maximum likelihood location for all the boxes in all layouts. In both (c) and (d), the predicted box locations are shifted towards a north-south plane in the center of the room.*

### 4.3.4 Dead-Reckoning Model with Calibration Errors

The dead-reckoning model simulates a moving observer storing an egocentric map of the box positions by constantly updating the heading direction with respect to north ($\alpha$) and estimating the distance traveled on each step ($d$). Here, we assume that the observer misestimates $\alpha$ and $d$ with a constant error (multiplicative calibration errors $\omega_\alpha$ and $\omega_d$ respectively). This leads to a cumulative error in the estimate of the participant's location. The box locations are assumed to be known correctly.

Initially, the position of the boxes is given, by definition, as follows, where $S_0^b$ is the starting distance of box $b$ and angle $\eta_0^b$ is its visual direction with respect to 'north' ($\frac{\pi}{2}$) as viewed from the start position (boxes are still visible here):

$$\eta_0^b = \frac{\pi}{2} - \text{atan2}(boxpos_x^b - pos_{0,x}, boxpos_y^b - pos_{0,y})$$

$$S_0^b = \sqrt{(boxpos_x^b - pos_{0,x})^2 + (boxpos_y^b - pos_{0,y})^2}$$

with $boxpos_x^b$ and $boxpos_y^b$ being the $x$- and $y$-coordinates of the $b^{th}$ box, $pos_{0,x}$ and $pos_{0,y}$ the Cyclopean point of the observer at the start position. At subsequent steps, when the boxes become obscured by the wall, the polar coordinates are calculated using the following equations:

$$\eta_n^b = \arctan\left(\frac{b}{a}\right)$$

$$S_n^b = \sqrt{a^2 + b^2}$$

with

$$a = \sin(\eta_{n-1}^b) \cdot S_{n-1}^b - \sin(\alpha \cdot \omega_\alpha) \cdot (d \cdot \omega_d)$$

$$b = \cos(\eta_{n-1}^b) \cdot S_{n-1}^b - \cos(\alpha \cdot \omega_\alpha) \cdot (d \cdot \omega_d).$$

We fitted this model to the data from Experiment 1 by varying the two free parameters, $\omega_\alpha$ and $\omega_d$, to give the minimum root-mean-square-error. (see Fig. S9).
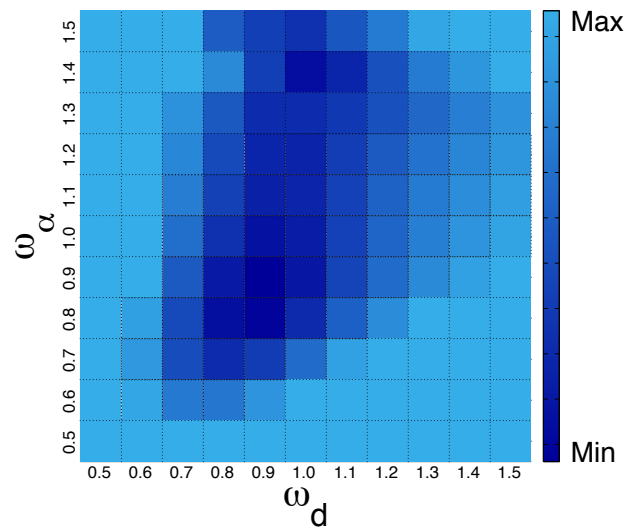
**Figure S9:** *Parameters in the dead-reckoning model. Using the data from all participants, this plots shows the different values of the root-mean-square error (RMSE) between data and model measured in degrees using different combinations of the model parameters $\omega_d$ (distance bias) and $\omega_\alpha$ (angular bias). Any combination with an RMSE larger than twice that of the ground truth model for the indirect condition of Experiment 1 ($RMSE_{lim} = 2 \cdot 16.6°$ ) has been set to that value ('Max'). Dark blue squares show the calibration errors with lower RMSE values, the lowest is at $(0.9, 0.9)$ with an RMSE value of $13.2°$ ('Min').*

### 4.3.5 Comparison between direct and indirect walking using dead-reckoning models
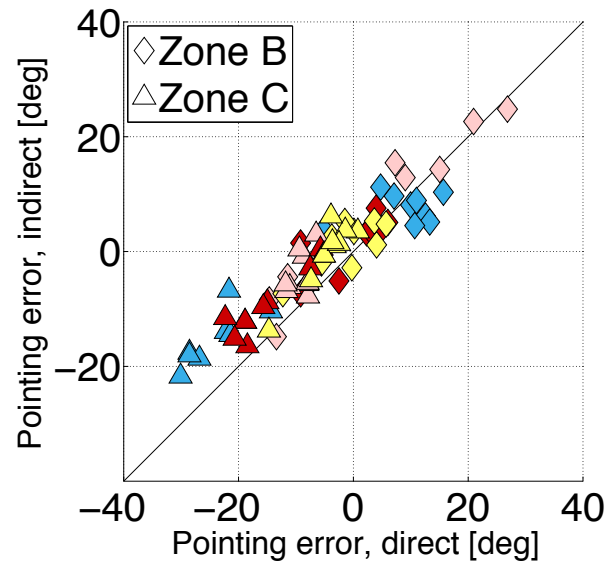


**Figure S10:** *Comparison of dead-reckoning predictions for direct and indirect walking. Error predicted by the dead-reckoning model for the indirect walking condition plotted against this value for direct walking trails. As in Figure 2d, each symbol is based on the mean data for 20 participants. The data for Zone C (triangles) is most informative as the length difference between the direct and indirect paths is most extreme in this case. Here, the errors for indirect walking are significantly more positive than the errors for direct walking (direct walking $M = 6.06, SD = 3.19$, indirect walking $M = 0.764, SD = 2.76$, $t(35) = 18.5, p < 0.001$) whereas the experimental data for these two conditions were not significantly different.*

### 4.3.6 Zero-mean noise

If, instead, we assume that the dead-reckoning noise has zero mean then there is no systematic effect on pointing, which we demonstrate as follows for an estimate of orientation. We added a normally distributed random error to the estimate of visual direction with respect to north on each step, $\eta_n^b$:

$$\eta_n^b = \arctan\left(\frac{b}{a}\right) + E$$

$$E = randn\left(0, \frac{\pi}{360}\right)$$

(3)

with the function $randn(\mu, \sigma)$ returning a random number drawn from a distribution with a standard deviation $\sigma$, and a mean $\mu$. $E$ is a random error added to the estimate of $\eta_n^b$, drawn from

a distribution with a standard deviation of $\sigma = \frac{\pi}{360}$ radians, and a mean of $\mu = 0$.
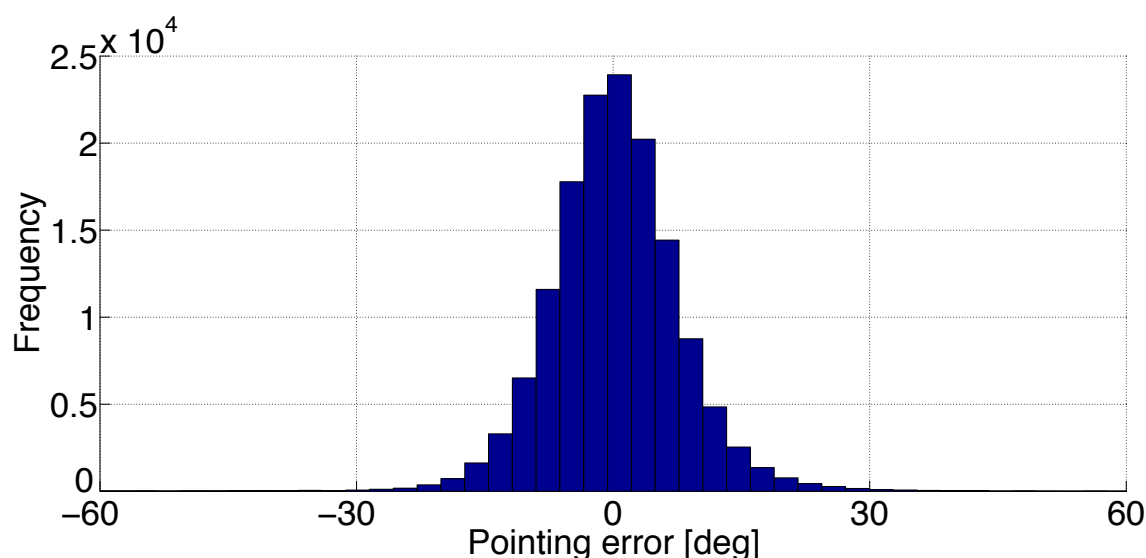


**Figure S11:** *Histogram of prediction errors calculated a 100 times for each box in each layout, using the walking trajectory of every participant tested in the indirect walking condition of Experiment 1 at pointing zones A, B, and C with a normally distributed random noise on the estimate of $\eta_n^b$. The mean of the pointing errors is not significantly different from zero.*

Using Eq. (3), predictions of pointing directions were calculated with a random additive noise on the direction of 'North'. Calculating the directions 100 times for each box in each layout, using the walking trajectory of every participant tested in the indirect walking condition of Experiment 1 at pointing zones A–C, a histogram of errors is plotted in Fig. S11.

### 4.3.7 Abathic Model

Fig. S12 shows psychophysical data from Johnston [44] that determine an abathic distance (shown by the cross, 82cm) and slope (0.23) relating estimated (or 'scaling') distance to physical viewing distance. In general, we can fit these two parameters to our pointing data (Figure 4a).
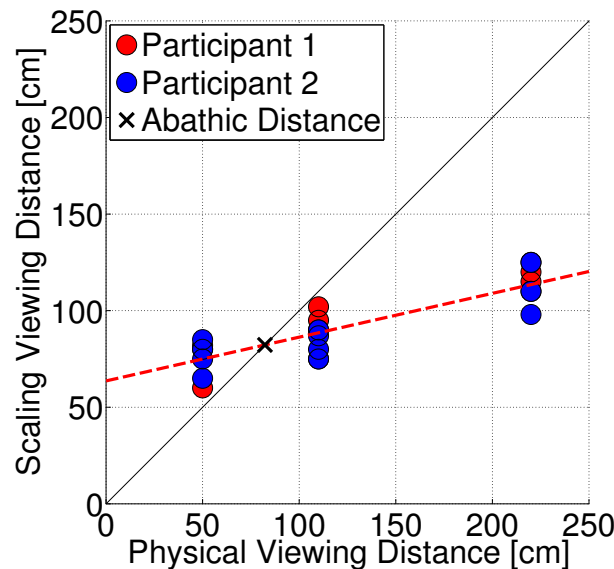


**Figure S12:** *Data re-plotted from Johnston's original paper [44] showing the relationship between true and estimated perceived egocentric distances. Note, that because the slope is less than 1 there is an underestimation of distances larger, and an overestimation of distances smaller than the abathic distance.*

The best fitting values are a slope of 1.03 and an intercept of 0.17 which corresponds to an abathic distance of $-5.66$. Note that the negative abathic distance is due to the slope being very close to 1. Thus, the misestimated egocentric distances of the boxes, $d_{est}^b$ is:

$$d_{est}^b = d_{true}^b * 1.03 + 0.17 \tag{4}$$

where $d_{true}^b$ is the true egocentric distance and $b = [1, \ldots, 4]$ the index for each box.

### 4.3.8 Comparison between abathic and dead-reckoning model predictions
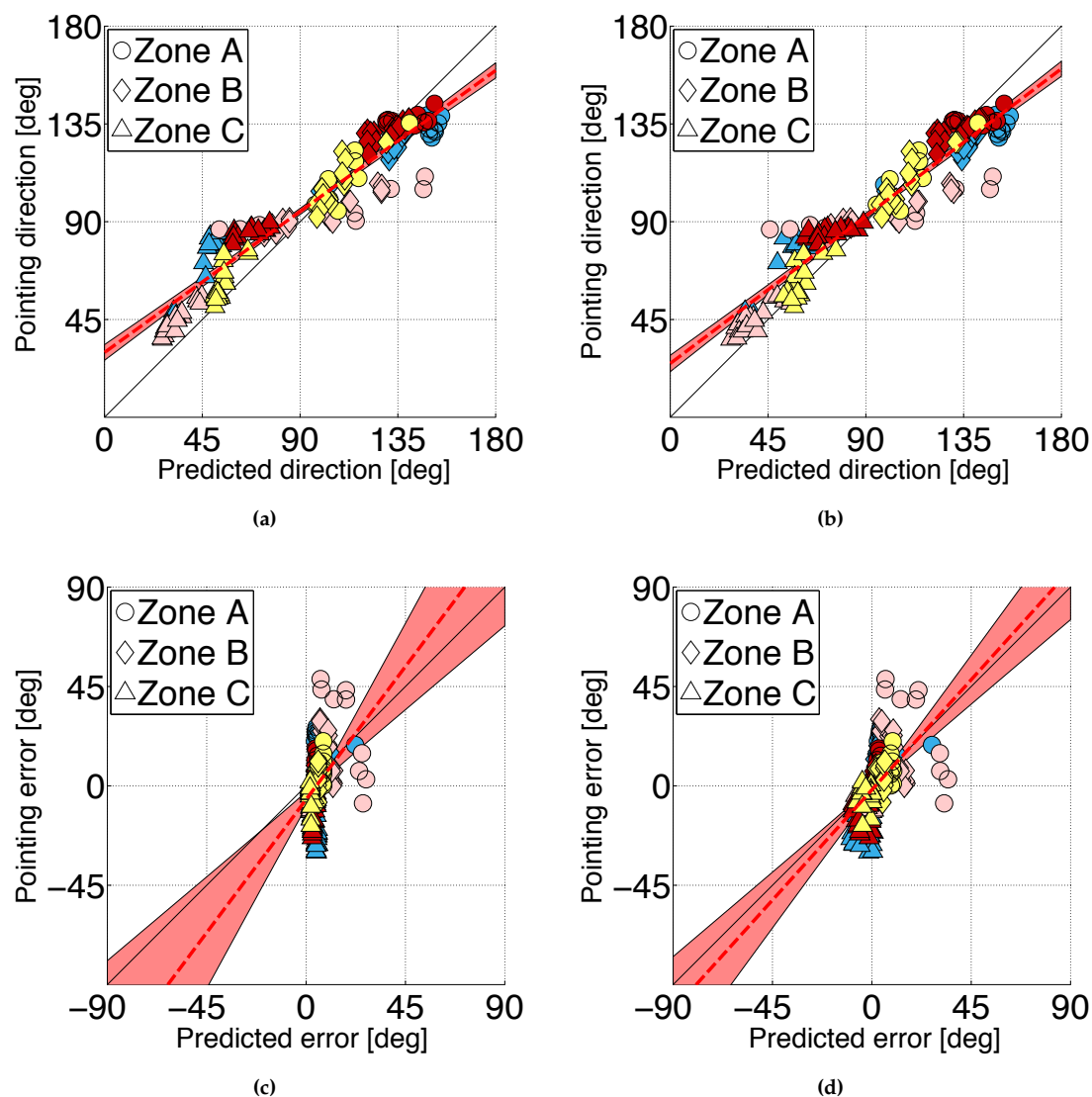


**Figure S13:** *Abathic and dead-reckoning models applied to the indirect walking condition in Experiment 1. (a) and (c) show pointing directions and errors for the abathic model. (b) and (d) show the same for the dead-reckoning model.*

| Participant | Root-Mean-Square Error [Degrees] | | |
| --- | --- | --- | --- |
| | Abathic Model | Projection Model | Dead-Reckoning Model |
| P1 | 11.55 | 7.65 | 16.11 |
| P2 | 21.06 | 18.73 | 27.54 |
| P3 | 10.33 | 7.89 | 11.68 |
| P4 | 13.34 | 10.93 | 13.92 |
| P5 | 19.28 | 17.57 | 23.74 |
| P6 | 8.43 | 11.21 | 8.66 |
| P7 | 12.97 | 11.68 | 12.66 |
| P8 | 23.92 | 19.76 | 22.56 |
| P9 | 12.98 | 9.57 | 14.56 |
| P10 | 21.39 | 14.45 | 21.16 |
| P11 | 19.61 | 15.72 | 18.15 |
| P12 | 9.36 | 17.80 | 16.23 |
| P13 | 16.79 | 11.87 | 14.85 |
| P14 | 14.36 | 10.66 | 16.78 |
| P15 | 10.09 | 13.15 | 15.75 |
| P16 | 12.57 | 10.46 | 11.27 |
| P17 | 9.69 | 13.09 | 12.05 |
| P18 | 10.31 | 13.65 | 14.74 |
| P19 | 16.40 | 11.11 | 15.74 |
| P20 | 13.75 | 11.12 | 15.47 |

**Table S1:** *Abathic, dead-reckoning and projection plane models compared. Gray-colored cells show lowest root-mean-square error amongst the 3 different models. ANOVA to test null hypothesis, that all the means across the 3 different models are drawn from the same distribution:null hypothesis cannot be rejected, with $\alpha = 0.05$, $[F_{2,57} = 3.00, p = 0.0579]$; right-tailed t-test to test whether the mean of errors of the abathic model ($M = 14.4, SD = 4.57$) is higher than the mean of the projection model ($M = 12.91, SD = 3.47$) (null hypothesis: the mean is not higher, alternative hypothesis: it is higher): $[t(19) = 1.78, p = 0.0457]$; right-tailed t-test to test whether mean errors of dead-reckoning model ($M = 16.2, SD = 4.58$) is higher than the mean error of projection model: $[t(19) = 4.69, p < 0.001]$*

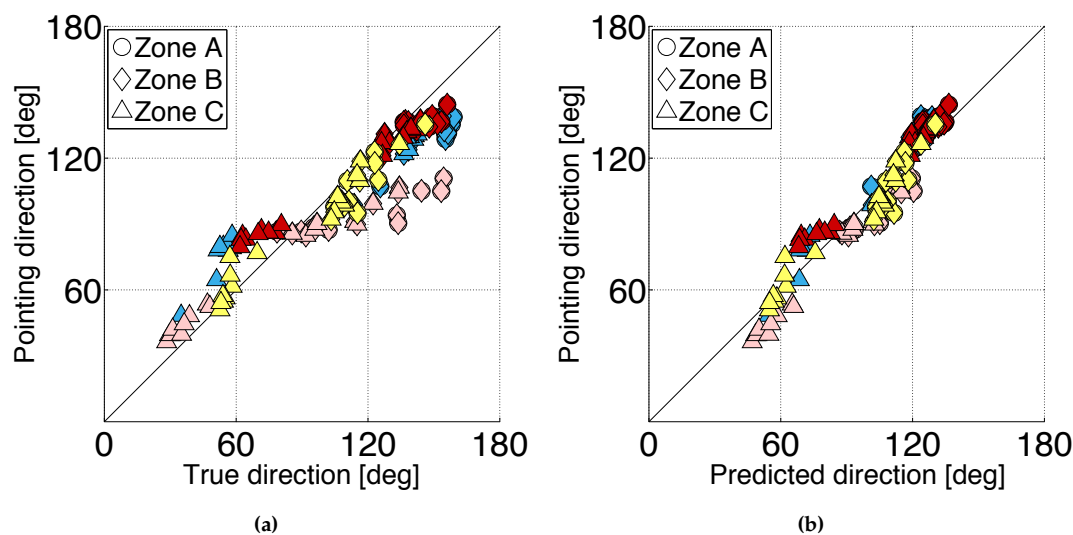### 4.3.9   Comparison between ground truth and projection model pointing directions



**Figure S14:** *Experiment 1 - actual and predicted pointing directions according to the projection plane model*
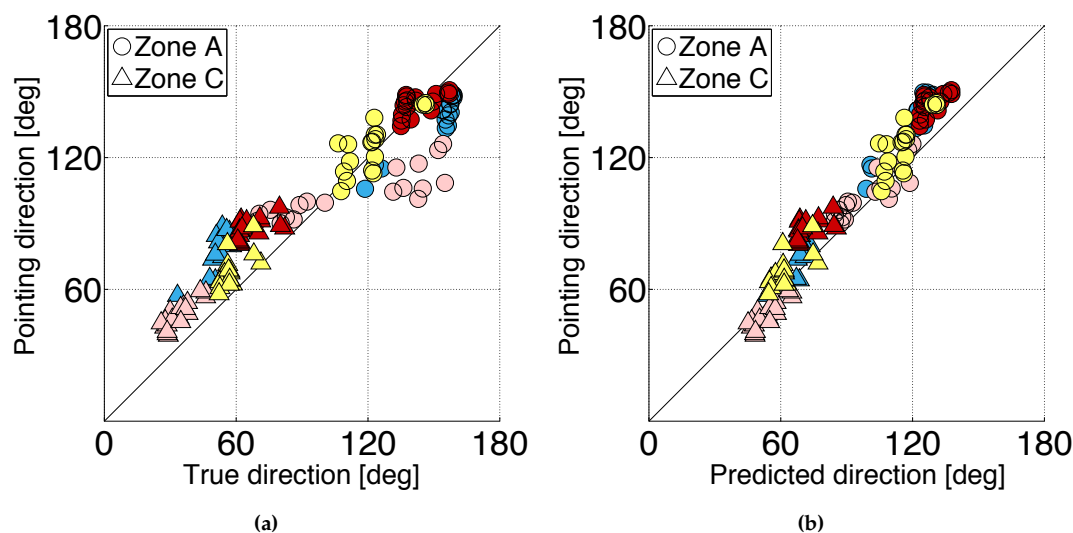


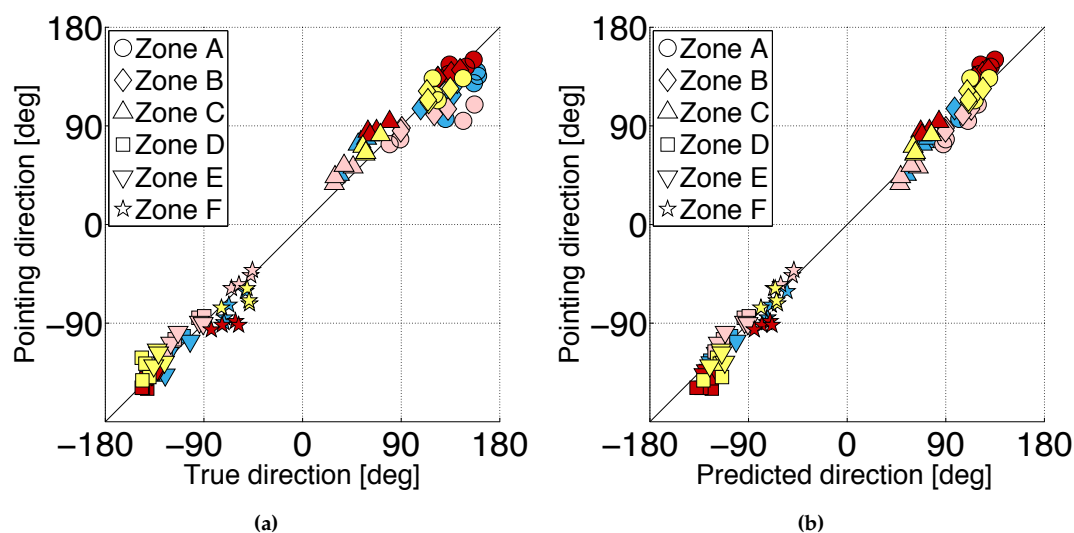**Figure S15:** *Experiment 2 - actual and predicted pointing directions according to the projection plane model*

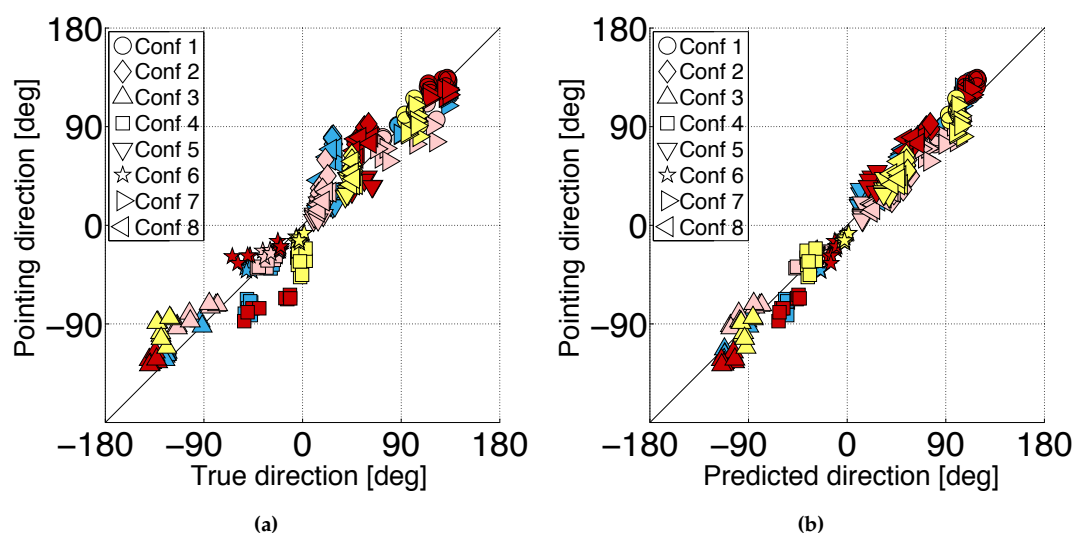**Figure S16:** *Experiment 3 - actual and predicted pointing directions according to the projection plane model*



**Figure S17:** *Experiment 4 - actual and predicted pointing directions according to the projection plane model*

### 4.3.10 Model comparisons using two model selection criteria, AIC and BIC

| Experiment | 1 | | 2 | | 3 | | 4 | |
|---|---|---|---|---|---|---|---|---|
| **Criteria** | Projection | Retrofit | Projection | Retrofit | Projection | Retrofit | Projection | Retrofit |
| AIC | 156.7 | 258.5 | 362.2 | 509.5 | 460.7 | 446.4 | 961.7 | 1177.7 |
| BIC | 158.3 | 298.1 | 364.5 | 566.4 | 463.3 | 510.5 | 964.9 | 1259.1 |
| # of Parameters | 0 | 24 | 0 | 24 | 0 | 24 | 0 | 24 |

**Table S2:** *A comparison of the fits of the projection plane model and the retrofit model using Akiake (AIC) and Bayesian (BIC) Information Criteria which penalize a model according to the number of free parameters it has.*

## 4.4. Raw data

Raw data and code to reproduce an example figure (Figure 8) are online:

https://github.com/jnyvng/PointingData .

Also, there is an interactive website showing the raw pointing data:

http://www.jennyvuong.net/dataWebsite/