

Psychophysics of Place Recognition

Hanspeter A. Mallot¹(✉), Stephan Lancier¹, and Marc Halfmann²

¹ Department of Biology, University of Tübingen, Tübingen, Germany
hanspeter.mallot@uni-tuebingen.de

² Leibniz-Institut für Wissensmedien, Tübingen, Germany

Abstract. Places are locations of special significance represented in spatial memory. Place recognition is a central task in spatial cognition that combines perception of local position information and ego-motion with working memories of adjacent places and long-term memory codes of the target place. In this paper, we examine the role of visual position information and place recognition and thus attempt to link spatial cognition to visual psychophysics. We present two experimental paradigms for assessing the visual processing involved in the formation of memory codes of place and the content of such memory codes. We also present a maximum likelihood model of place recognition from distant landmarks allowing detailed quantitative testing of the general assumptions. We conclude that place recognition is based on a visual working memory containing raw “snapshot” information as well as local depth maps of surrounding landmark objects.

1 Introduction

1.1 Place Recognition

Place vs. Location. The concept of place, i.e. of locations of special significance, is central to the understanding of spatial cognition. Origins of spatial memory in the animal kingdom are associated with a life-style known as central-place foraging (Papi 1992) in which animals keep and remember a “home” location from which they make excursions for feeding. Indeed, the representation of this central place may be the simplest case of a spatial long-term memory in the animal kingdom. Finding back to the central place can be based on various mechanisms including search, laid-out trails (chiton), path integration (ants, spiders, honeybees), or landmark-based matching (digger wasp, honey-bee). In any case, the animal will have to know when the home is reached, in which case some change in behavior will occur, e.g. the animal will decide to stop moving or to start a final search routine.

In rodents and other mammals, multiple places or location in general are thought to be represented in the activity of hippocampal place cells. For example, Wilson and McNaughton (1993) demonstrate that the current location of

Supported by the German Federal Ministry of Education and Research within the Tübingen Bernstein Center for Computational Neuroscience (Grant 01GQ1002A).

a rat in a maze can be reconstructed from a population of ongoing place-cell recordings, given that the firing fields of these neurons had been determined in a previous measurement. Still, firing fields do not simply pave environmental space in a homogeneous way. Rather, density and overlap of place cell firing fields is increased at places which have an increased significance for the animal. For example, Hollup et al. (2001) showed that in rats trained to find a hidden platform in an annular watermaze, more firing fields are found in the vicinity of the platform than in mid-water. Firing field density changes as the platform is relocated.

In primates, place recognition also involves other brain regions, including among others the parahippocampal place area (Epstein and Kanwisher 1998, Epstein 2008). Places represented in the parahippocampal place area are discrete entities characterized not just by their location (expressed e.g. by geometric coordinates) but by invariant features such as landmark objects or the overall geometrical layout of a scene.

Cognitive Graphs. Cognitive models of spatial behavior on the navigational or way-finding scale are often based on places as a central data format or “spatial primitive”. For example, the base level of Kuipers’ (2000) “spatial semantic hierarchy” is formed by places which are recognized and approached by the minimization of some measure of perceptual distance between the place and the agent’s current location. The place representations are connected by action links allowing the agent to travel from one place to the next. Tolman’s “means-ends-field” (Tolman 1932) is also a graph-like structure in which the nodes are states of the animal which may include the recognition of being at a particular place as well as goals which the animal is currently pursuing. Again, the states are linked by “means-ends-relations” allowing to plan state transitions. A developmental argument for the relevance of places as a building block in spatial memory has been presented by Siegel and White (1975). Graph approaches underly a large part of the wayfinding literature in which routes are generally considered chains of recognized places and actions, see for example O’Keefe and Nadel (1978), Gillner and Mallot (1998), and Hartley et al. (2003).

Despite the central role of “places” in many representational formats of space, other structures with similar roles may exist in spatial memory. One possibility is the oriented “view” visible from a location. View-specific neurons have been found in the primate hippocampus by Rolls et al. (1998) and have been used as nodes for cognitive graphs e.g. by Schölkopf and Mallot (1995) and Gaussier et al. (2002). Spatial graphs may also be composed of regional nodes representing groups of places in a hierarchical scheme (Wiener and Mallot 2003) or patch maps including a local reference frame (Meilinger 2008). Views, places, and regions differ in the granularity of spatial representations. In the experiments reported here, the extension of a place is mostly treated as an uncertainty, quantified by a confusion area, i.e. the statistical error ellipse of the place judgments. A more comprehensive theory of place recognition should probably treat the extension as a property of a mentally represented place.

Views, as well as places, as elements of a spatial ontology are associated with a geometrical location in the sense that they are perceived when the agent is located at or looking from this location. The situation is different for landmarks and boundaries, which may also be elements of the spatial graph and associated with specific locations, but which need not be reachable for the navigating agent. In this paper, we consider reachable places which are encoded in memory during an actual visit at this place and are recognized during subsequent encounters. Landmarks and boundaries will show up only as descriptors of places, not as nodes of the cognitive graph.

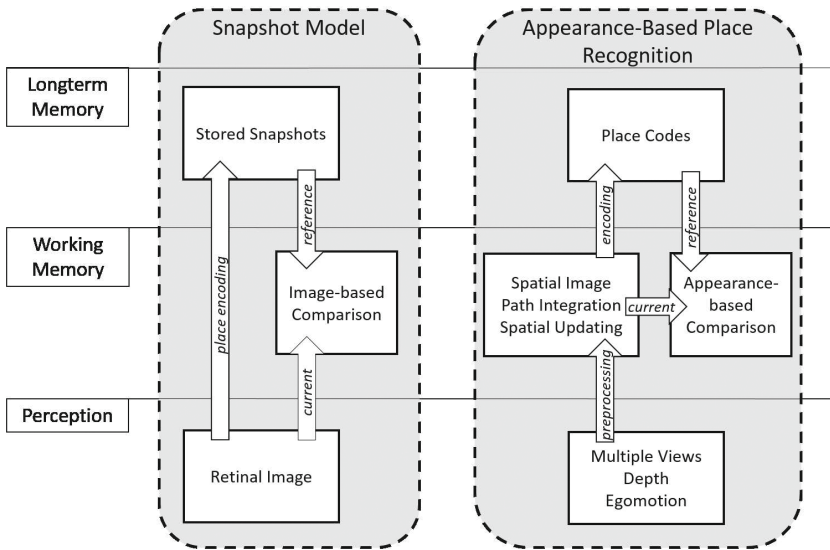


Fig. 1. Two approaches to place recognition

1.2 Models and Mechanisms

Figure 1 summarizes two basic approaches to place recognition. In the snapshot approach suggested by Cartwright and Collett (1982), left part of Fig. 1, the memory code for a particular place is closely related to the retinal image visible at the encoded place. Since this image may change substantially with the illumination, the time of day or year etc. (Zeil et al. 2003), some preprocessing is essential to yield a sufficient level of invariance. Such preprocessing of the retinal image has indeed been suggested by Cartwright and Collett (1982), who assume that the retinal image passes an edge detector before being stored as a snapshot. Other preprocessing operations include detection of “landmark” pixels (Lambrinos and Möller 1997), sky-line detection (Graham and Cheng 2009; Basten and Mallot 2010), average egocentric position of landmark objects (“center-of-gravity”, O’Keefe 1991), local distance map derived from motion parallax (Dittmar et al. 2010), etc. In any case, the actual recognition step

is carried out as a comparison or matching operation between the stored and a current snapshot, both preprocessed in the same way. Algorithms for this comparison operation include feature correspondence search (Cartwright and Collett 1982), pixel-based root-mean-square minimization with and without image warping (Franz et al. 1998) etc. For review see Möller and Vardy (2006).

In human spatial cognition place recognition can be based on snapshot-like mechanisms (Gillner et al. 2008). However, it is quite clear that under normal circumstances, extra-retinal information provided by one or several working memory stages will also play a role. These working memories contain information from outside the current field of view and are updated upon observer motion (Loomis et al. 2013; Burgess and Hitch 2005; Schindler and Bartels 2013; Röhrich et al. 2014). They provide a description of the local spatial layout which may be conceptualized as a local map or environment model surrounding the subject. Behavioral experiments with configurations of isolated landmark objects show that local charts of such landmark objects play a role in place recognition (Waller et al. 2000; Pickup et al. 2013). Here, we will use the term “spatial appearance” to characterize the spatial working memory of a place. Place codes for long-term memory (LTM) will then be derived from the spatial appearance and recognition will be based on comparisons of such LTM codes with the current spatial appearance (right part of Fig. 1).

While the idea of appearance-based place recognition does not seem to be particularly controversial, it allows for a number of systematic questions that may be used to structure the psychophysics of place recognition. These questions are:

1. Depth of processing: which image processing steps are needed to extract the place code from the raw retinal image? Possibilities include various early vision operations (edge detection, parallax) as well as the recognition of more abstract landmark structures such as the sky-line or recognizable objects.
2. Structure of the spatial working memory: which information is maintained and used for place recognition? Possibilities include panoramic views, collections of local views, objects localized in a local, egocentric but two-dimensional map, three-dimensional spatial layouts, etc.
3. Mechanisms of comparison and decision making. These will depend on the structure of the working memory and may include pointwise snapshot matching, comparisons of local maps, identification of landmark objects, etc.
4. Structure of long-term memories of places. This includes the role of local metric charts, geometric layout of the surrounding scene, and context from larger representations of space.

1.3 Experimental Procedures

The classical place recognition experiment in animals has been developed by Kruyt and Tinbergen (1938) in a study on digger wasps returning repeatedly to each of a number of borrows in which a larva awaits feeding. In this case, the place is selected by the animal and no learning or specific training is required

during the experiment. Tinbergen and Kruyt (1938) marked a borrow by a circle of pine cones and displaced this circle after the digger wasp had left from the borrow. When returning, the wasp would search in the center of the displaced circle, indicating that the cones were used as landmarks.

In the Morris water maze task for rodents (Morris 1981) a goal location is rewarded by the fact that a rat can rest from swimming when it reaches the submerged platform in a water basin. Place learning in the water maze assesses both the recognition performance and the learning speed. The experiments can last for extended time, suggesting that a long-term memory of the place of the submerged platform is built and tested. For human subjects, various versions of the Morris water maze have been realized in virtual reality. For example, Hamilton and Sutherland (1999) used a virtual pool surrounded by two sets of landmarks (A, B) to study blocking in landmark learning. When trained with both sets together (A + B), subjects performed well even after one set was removed. However, when trained with one set only (A) and given the additional set B later, removal of set A lead to performance loss, indicating that the landmarks B had not been learned in this situation. Hort et al. (2007) adapted the water-maze paradigm to humans using a circular arena of 2.9 m diameter with landmarks projected variably by a beamer. This setup was used to demonstrate deficits in spatial cognition associated with early stages of dementia.

In this paper, we will study place recognition by two different experimental tasks. In the first one, “return-to-cued-location” (Gillner et al. 2008), a subject in a virtual environment is presented with the visual surround visible from a goal location. The virtual viewpoint tracker is then switched to a starting point from which the goal has to be approached by interactive navigation in the virtual environment. The task is similar to the tasks used in the geometric module literature (Cheng et al. 2013), but goal places can occur anywhere in the maze. Since the memory of the goal location is built during a brief inspection period prior to the performance, the “return-to-cued-location” task addresses a working memory of place. In the second paradigm, “incidental place learning” subjects are required to navigate to a goal from different starting points, using routes which all share a central crossing point. In the learning phase, this crossing point is always passed, but never mentioned as a special point to remember. In the test phase, however, subjects are explicitly instructed to navigate to the central crossing point. This paradigm has the advantage that subjects have to discover the central point themselves. It emphasizes spatial long-term memories.

The problem of place recognition is closely related to place learning, and experimental paradigms will generally involve both performances. It is important to note, however, that different learning schemes may lead to different place representations. One respect in which these representations may differ is their characterization as long-term or working memories. A second respect is place selection which is arbitrary in supervised schemes such as the Morris water maze or the walk-to-cues-location paradigm but may be influenced by the availability of landmarks etc. in free place choice.

2 Depth of Processing

Place recognition from visual cues involves the standard processes of early vision, including among others the detection of image features and depth, the understanding of scenes, and the recognition of objects. Here we use stereoscopic dynamic random dots to study the role of pure depth information in place recognition. Results indicate that place recognition can be based on pure depth information and (at least in our experimental environment) is not substantially improved by cues from other visual sub-modalities such as texture or localized objects (room corners).

2.1 Local Position Information

The recognition of places is generally thought to rely on a combination of landmark cues visible from the target place and spatial context such as traveled distances from neighboring places (e.g., O’Keefe and Nadel 1978). For the landmark component, various types of “local position information” can be extracted from the visual input and have been shown to play a role in place recognition. These types include barely processed “snapshots” (for review, see Gillner et al. 2008) as well as visual information requiring higher amounts of image processing such as landmark configurations (see next section), room geometry and three-dimensional spatial layout (Cheng et al. 2013; Epstein 2008), or identified landmark objects (Janzen and van Turenout 2004). Visual depth, i.e. the perceived distance to objects of the surrounding scene, is relevant for a number of these cues, especially if indoor-environments are considered. Here we use psychophysical approaches from the study of early visual processes (stereopsis, motion parallax) to investigate the role of perceived depth in place recognition (Halfmann 2016).

2.2 Methods

Subjects and Procedure. 40 students from the University of Tübingen passed a simple test for stereo vision and participated in this study. The experiments were carried out in a virtual environment simulating a kite-shaped room with edged or rounded corners. In the “return-to-cued-location task” (Gillner et al. 2008), participants were placed at one of three goal locations in the kite-shaped room. In the following inspection phase subjects studied the local appearance of the room by looking around and performing small translational movements. They were then set back to a start position and used a joy-stick to return to the goal. After indicating goal recognition by the button hit, subjects were moved to the correct goal position, and the next trial started from there. In all, twelve decisions were recorded per subject and condition, i.e. two cycles of all six possible transitions between the three goal locations. In the results reported here, the virtual environment was presented with an Oculus-Rift stereoscopic head-mounted display (HMD), but controls with a mirror stereoscope and monocular viewing were also performed. In addition to the stereo disparities presented on

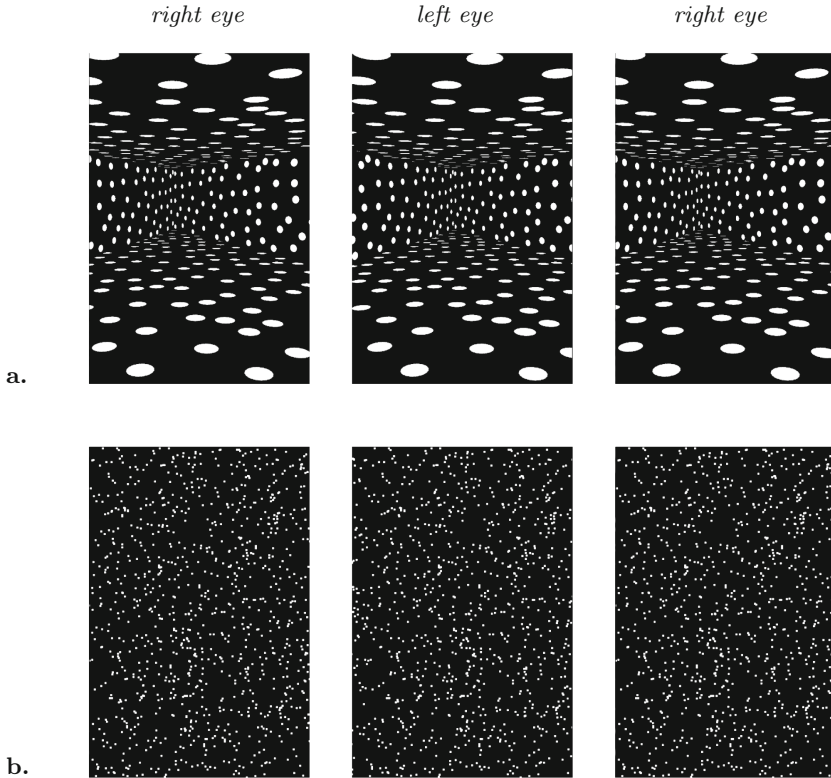


Fig. 2. Sample view of the kite-shaped room arranged for free stereoscopic viewing. For crossed fusion use leftmost columns, for uncrossed fusion use rightmost columns. **a.** Texture condition. **b.** Dot condition (sample frame of the dynamic random dot display). Both stereograms show the room with edged corners. Note that the texture condition gives a much better stereoscopic impression than the dot condition. However, in the actual experiment, motion parallax was present as an additional cue, leading to a clear perception of the room layout.

the stereoscope, the HMD setup provided a higher level of immersion including closed-loop movements of the head and body that might lead to better perception of structure-from-motion.

Stimuli and Conditions. Two factors, “visual cues” and “room shape”, were varied in a full factorial design. In the *cue-condition* “*texture*”, rooms were defined by a texture of large spots (about 10 cm diameter in the virtual environment) pasted to the room walls, floor, and ceiling as a wallpaper. This texture provided stereo disparity, motion parallax upon observer motion, texture gradients and information about room corners (Fig. 2a). In the *cue-condition* “*dots*”,

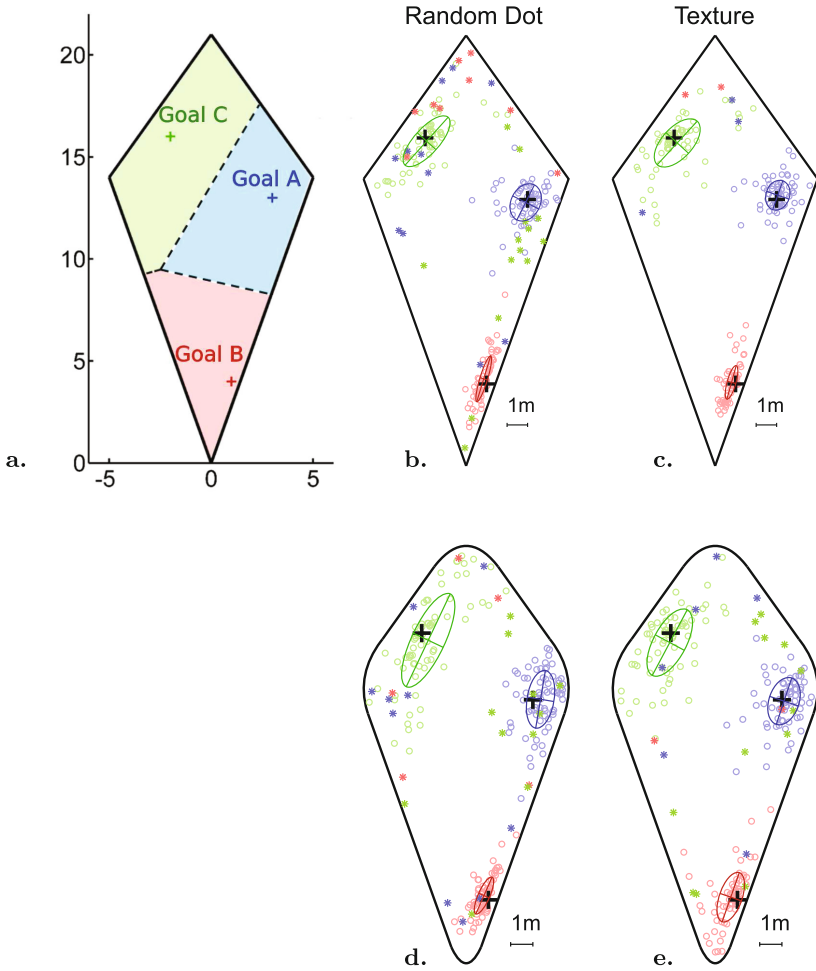


Fig. 3. Scatter plots of the decision points from 240 decisions (20 subjects \times 12 decisions per subject). **a.** Layout of the kite-shaped room with goal locations A, B, C, and nearest-neighbor cells. Dimensions in meters. **b., c.** Edged corner room, **d., e.** rounded corner room. Dot colors indicate goal positions A, B, C. Tokens indicate: + true goal location, \circ decision points within goal region (“correct decision”), * decision point outside goal region (“qualitative error”). Error ellipses are calculated over the within-region decisions only and reflect one standard deviation. (Color figure online)

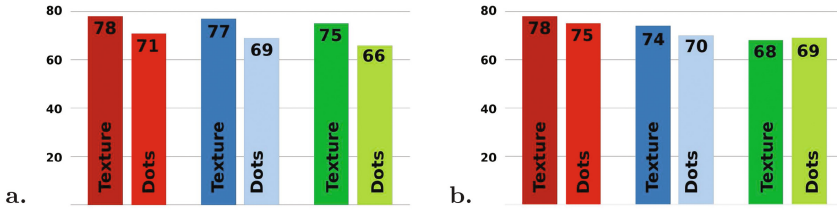


Fig. 4. Absolute numbers of correct decisions (decisions inside goal region) out of a total of 80 decisions per target (accumulated over all subjects). **a.** Edged corner room, **b.** rounded corner room. Colors indicate goal locations A, B, C. Performance above chance level defined by the relative area of Voronoi cell is highly significant for all cases. (Color figure online)

surfaces were defined by dynamic random dots (Sperling et al. 1989) uniformly distributed in the image plane and with a limited lifetime varying between 100 and 200 ms. Outdated dots or dots leaving the field of view were continuously replaced so that the dot distribution on the screen was kept uniform. The dots provided stereo disparities, a small amount of motion parallax (during dot lifetime), but no texture gradients (see Sperling et al. 1989). Room corners might have been inferred from the depth information, but not from the dot distribution itself (Fig. 2b). The cue conditions were performed in a blocked, within-subject design (texture condition first).

Even if the example stimulus of Fig. 2b is properly fused, the structure of the room is barely visible. In the experimental setup, however, the dots would start to move as soon as the observer changes his or her viewpoint. In this situation, the three-dimensional structure of the room becomes much clearer, since motion parallax can be used.

We used two *shape-conditions* “edged”, and “rounded”, as shown in Fig. 3 (between subjects factor). These conditions were included to test the hypothesis that the better defined corners in the “edged” condition provide better landmark information than the rounded corners in the “rounded” condition, predicting a superior performance in the “edged” condition.

2.3 Results

Figure 3 shows the decision points in the four conditions, accumulated over all subjects. Decision points scatter about the goal positions with a moderate variance, and variance is not substantially different in the four conditions. We also find a fair number of “qualitative errors” in which the subjects choose a place closer to one of the non-goals than to the current goal. The respective nearest-neighbor cells (Voronoi tessellation around goal points) are also indicated in Fig. 3a. These errors are equivalent to the “rotation errors” discussed in the geometric-module literature (see Cheng et al. 2013 for review). Figure 4 shows the number of correct decisions for the various conditions, again accumulated over all subjects. Note that the numbers given there are absolute counts out of

80 trials and do therefore not carry error bars. If subjects would ignore the visual information, the chance level for choosing a decision point in the correct Voronoi cell would be about 33% compared to an average recorded performance rate of about 91% shown in Fig. 4. A binomial test with chance level as null hypothesis reveals high significance ($p < 0.001$) in all cases. No significant differences between conditions were found.

A comparison with the stereoscopic and monocular viewing conditions (data not presented in this paper) shows similar results. Performance is well above chance even for the monocular condition, albeit slightly poorer than in the HMD-data reported here.

2.4 Discussion

The results indicate that subjects can use pure depth information as is provided by dynamic random dots to recognize places in a room. Additional texture cues providing more reliable depth information seem to lead to some improvement, which, however, is not statistically significant. This is even more surprising since texture cues provide still another cue for place recognition, i.e. snapshot matching. Indeed, since the texture was “painted to the wall”, the subjects might have tried to remember the pattern of black and white wall patches appearing at each goal location and try to match it to their memory when they return. If they did use this strategy, it did not lead to a substantial improvement in performance. The sharpness of the corners of the room (“edged” vs. “rounded” conditions) do not seem to play an important role in self-localization, indicating that subjects rely more on the distances to walls than to the corners. Overall, the results fit nicely to the idea that places are represented by a local map of the environment which is updated as the subject moves around (Byrne et al. 2007; Loomis et al. 2013; Röhrich et al. 2014).

3 Place Recognition from Distant Landmarks

In this section, we present experimental data and a probabilistic model of place recognition from a configuration of distant landmarks surrounding a goal. The model assumes that landmark positions are perceived with hyperbolic distance compression and added noise, depending on current observer position. Position-dependent recognition rate is modeled as the likelihood of perceiving the expected (stored) landmark configuration from each position. The model reproduces key features of experimental results including a systematic localization bias towards the most distant landmark, the shape and orientation of the error ellipses, and effects of approach direction. We conclude that place recognition is based on a comparison between a place code (landmark distance and angles) and a working memory of surrounding space suffering from systematic depth distortions and distance-dependent drop in resolution.

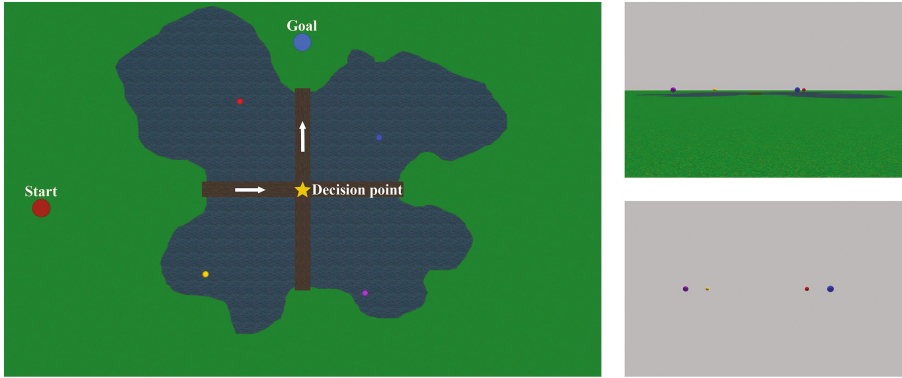


Fig. 5. Experimental setup for Experiment 2. *Left:* Aerial view of pond and plus-shaped bridge with a start and a goal location. In the experiments, four start and goal locations close to the ends of the bridge arms were used in all possible combinations involving a left or right turn at the decision point (bridge center). Four landmark objects can be seen in the four quadrants defined by the bridge. *Top right:* Subjects' view during learning phase. Note the landmark objects hovering above the pond. *Bottom right:* Subject's view during test phase. Only the landmark objects remain visible while the pond and bridge are covered by fog.

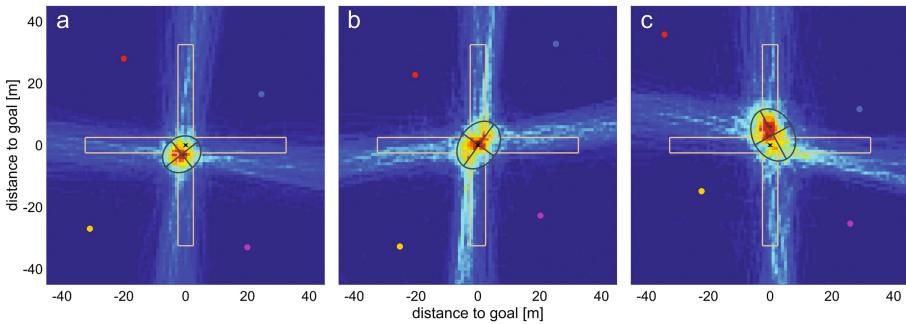


Fig. 6. Position choices for three landmark configurations. The landmarks are shown with their actual position and color. **a.** Standard configuration (20 subjects, 954 decisions), **b.** Parallelogram configuration (16 subjects, 761 decision), **c.** Peaked configuration (16 subjects, 754 decisions). The error ellipses are displaced from the goal (control and peaked condition) and elongated in the direction of the most distant landmark. See Lancier (2016).

3.1 Summary of Experimental Data

The accuracy of the place recognition in an open environment comprising four distant, distinguishable landmarks was studied in a behavioral experiment with human subjects navigating a virtual environment (Fig. 5). The environment included a plus-shaped bridge crossing a pond and four colored spheres hovering in mid-air above the pond, one in each quadrant defined by the bridge

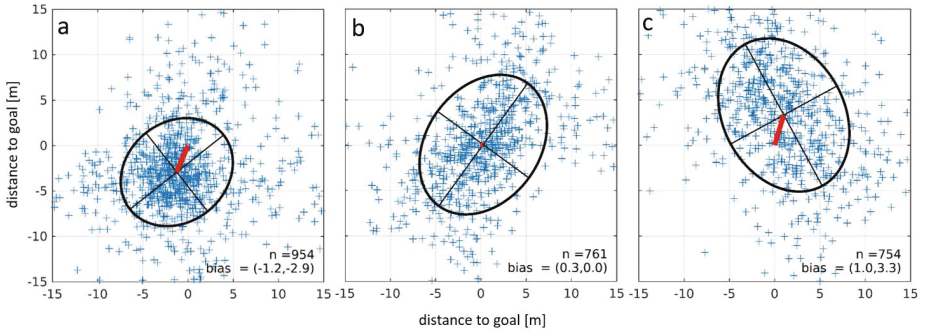


Fig. 7. Detailed decision points for the data appearing in Fig. 6. **a.** Standard configuration (20 subjects, 954 decisions), **b.** Parallelogram configuration (16 subjects, 761 decisions), **c.** Peaked configuration (16 subjects, 754 decisions). The read bar marks the mean deviation from the bridge center (“bias”); it was significantly different from zero for the standard and peaked conditions (Hotelling’s T-Square test).

arms. Subjects started at one bridge entry and had to find a goal that involved either a left or a right turn at the bridge center (“decision point”). All possible starting points and turn directions were used. In the test phase, bridge, pond, and goals were rendered invisible by simulated ground fog and the subjects were asked to navigate to the now invisible center of the bridge and indicate place recognition by button hit. This performance was based essentially on the four landmarks which remained visible at all times. In order to prevent subjects from using path integration, the starting points at each of the four bridge entries were varied using a random positional scatter. Experimental results are summarized in Fig. 6 (Lancier 2016). For the model, the following constraints can be derived:

1. Decision points show both a systematic bias and a statistical error. The systematic bias as well as the major axis of the error ellipses point roughly in the direction of the most distant landmark (Figs. 6 and 7).
2. If a point-symmetric configuration of landmarks is used, the systematic bias goes away (Figs. 6b and 7b).
3. If the landmark sizes, and therefore the perceived landmark distances, are manipulated between training and test session, decision points are shifted towards down-scaled landmarks and away from the up-scaled ones (data not shown). I.e. subjects try to adjust remembered and perceived distances.

3.2 Model

In a world coordinate system centered around the target point (the center of the bridge), the landmark positions are denoted by \mathbf{l}_i , $i = 1, \dots, 4$. Let \mathbf{x} denote the current observer position. The true landmark vectors from the current observer position are $\mathbf{m}_i = \mathbf{l}_i - \mathbf{x}$. We assume that these positions are represented in

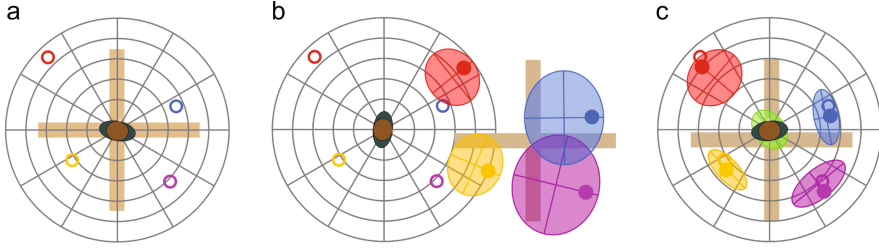


Fig. 8. Place recognition model. **a.** Layout with bridge (shown in light brown) and four landmarks arranged as in the “peaked” condition (Fig. 6c), shown as open circles. The polar grid symbolizes the egocentric landmark memory of an observer inspecting the bridge center. In this grid, the landmark positions are stored as a place code. **b.** Approaching observer with place code (open colored circles). Solid colored disks: true landmark positions; transparent ellipses: distribution of landmark measurement according to Eq. 2. Note the displacements of the distributions relative to the true landmark positions, which reflects the assumed hyperbolic distance compression. **c.** Probabilistic match of place code and observed landmark positions. The green distribution in the center is the joint likelihood $p(\mathbf{l}_1, \dots, p(\mathbf{l}_4|\mathbf{x})$ from Eq. 4. (Color figure online)

an egocentric coordinate system with allocentric orientation. This allocentric orientation can be provided by the overall orientation of the bridge and landmark configuration (Fig. 8).

In order to model the systematic bias, we will need to assume that the actual perceived landmark distance is not veridical but hyperbolically compressed according the equation

$$\boldsymbol{\mu}_i = \frac{A}{A + \|\mathbf{m}_i\|} \mathbf{m}_i \quad (1)$$

(Gilinsky 1951). A is a constant set to 60 m in our simulations. This compression does not affect the stored landmark position which is assumed to be derived from triangulation and spatial updating processes and may therefore be assumed veridical. Indeed, Philbeck and Loomis (1997) demonstrate that the distance walked to a visually presented target in the “walking-without-vision” task is not affected by the hyperbolic compression reported by Gilinsky (1951). The stored place code is therefore given by the true landmark positions \mathbf{l}_i .

Consider the probability of perceiving a landmark i at a position \mathbf{m}_i , given that the current observation position is \mathbf{x} . This measurement \mathbf{m}_i is given in a Cartesian, egocentric coordinate system oriented to some allocentric “North” orientation. It comprises information about the perceived egocentric distance (with hyperbolic compression) and allocentric bearing, i.e. bearing with respect to a reference direction defined, for example, by the overall orientation of the virtual environment. The probability density function is assumed to be

$$p(\mathbf{m}_i|\mathbf{x}) = \phi(\mathbf{m}_i; \boldsymbol{\mu}_i(\mathbf{x}), \Sigma_i(\mathbf{x})), \quad (2)$$

i.e. the two-dimensional normal distribution with mean $\boldsymbol{\mu}_i$ and covariance matrix Σ_i . Note that both mean and covariance depend on the current observer position \mathbf{x} . For the mean, we have specified this dependence in Eq. 1 above. The covariance matrix Σ_i will have an eigenvector in the direction $(\mathbf{l}_i - \mathbf{x})$, i.e. the depth direction from the current view-point to the true landmark position, and an orthogonal one in the width direction. Denoting the local bearing of the i -th landmark by ϕ_i , ($(\cos \phi_i, \sin \phi_i) = (\mathbf{l}_i - \mathbf{x}) / \|\mathbf{l}_i - \mathbf{x}\|$), we obtain:

$$\Sigma_i(\mathbf{x}) = \begin{pmatrix} \cos \phi_i & -\sin \phi_i \\ \sin \phi_i & \cos \phi_i \end{pmatrix} \begin{pmatrix} \sigma_{id}^2 & 0 \\ 0 & \sigma_{iw}^2 \end{pmatrix} \begin{pmatrix} \cos \phi_i & \sin \phi_i \\ -\sin \phi_i & \cos \phi_i \end{pmatrix}. \quad (3)$$

The eigenvalues in the distance and width directions are assumed to scale with distance according to $\sigma_{id}(\mathbf{x}) = 0.01 \|\mathbf{l}_i - \mathbf{x}\|^2$ and $\sigma_{iw}(\mathbf{x}) = 0.3 \|\mathbf{l}_i - \mathbf{x}\|$. Thus, the angular error of perceived landmark bearing does not depend on viewing distance. For small distances the angular errors are larger than the depth errors ($\sigma_{iw} > \sigma_{id}$) as is necessary to reproduce the shape of the experimental distributions. This may reflect the fact that inter-landmark angles have to be inferred from multiple views and are therefore more error-prone than the distance estimates.

As the observer moves, the probability densities $p(\mathbf{m}_i | \mathbf{x})$ will be shifted to their new bearing and (hyperbolically compressed) distance. In addition, they will be rotated to keep the principle axis associated with σ_{id} aligned with the landmark bearing. The place code for the goal position $\mathbf{x} = 0$ will be $\{\mathbf{l}_i, i = 1, \dots, 4\}$. The probability of measuring this place-code, given that the observer is actually at \mathbf{x} , is obtained by substituting $\mathbf{m} = \mathbf{l}_i$ in Eq. 2 and taking the product over all four landmarks:

$$p(\mathbf{l}_1, \dots, \mathbf{l}_4 | \mathbf{x}) = \prod_{i=1}^4 \phi(\mathbf{l}_i; \boldsymbol{\mu}_i(\mathbf{x}), \Sigma_i(\mathbf{x})). \quad (4)$$

The function $LL(\mathbf{x}) := \log p(\mathbf{l}_1, \dots, \mathbf{l}_4 | \mathbf{x})$ is plotted as the model prediction in Fig. 9 for the error distributions for the three landmark configurations appearing in Figs. 6 and 7.

Note that the likelihood function $p(\mathbf{l}_1, \dots, \mathbf{l}_4 | \mathbf{x})$ will always take its maximum at $\mathbf{x} = 0$ if we omit the hyperbolic distance compression (Eq. 1). In this case, we have $\boldsymbol{\mu}_i = \mathbf{l}_i - \mathbf{x}$ and the product in Eq. 4 is taken over four Gaussians all of which are centered at $\mathbf{x} = 0$. The systematic bias found in our experiments cannot be explained in this case.

The simulations of Fig. 9 are in good quantitative agreement with the experimental results appearing in Figs. 6 and 7. In particular, they reproduce the bias towards the most distant landmark in the standard and peaked configurations, and the orientation of the error distributions. A quantitative test of the model was obtained with the directional statistics of the decision points appearing in Fig. 10. Each decision point judgment was transformed into a unit vector and counted in a circular histogram. Figure 10 shows the resulting distributions together with the landmark bearings (colored circles) and the direction of the

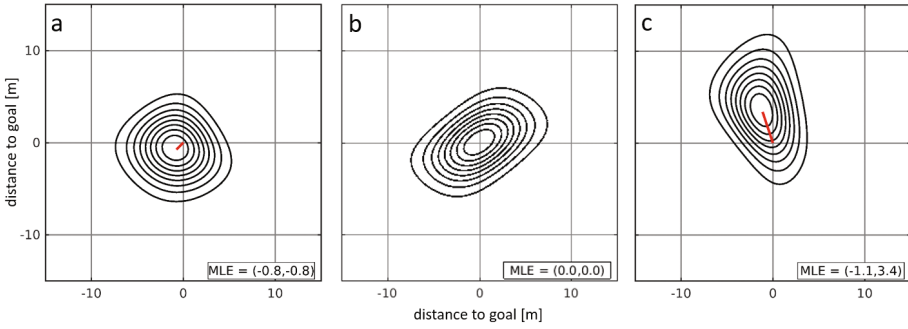


Fig. 9. Likelihood function (Eq. 4). **a.** Control condition, **b.** Parallelogram condition **c.** peaked condition. The red bar indicates the maximum likelihood estimator (MLE) for the bias. The direction and roughly also the length of the predicted biases agree with the experimental results (Fig. 7). Note that the likelihood distributions in **b.** and **c.** also show the elongation towards the most distant landmark. (Color figure online)

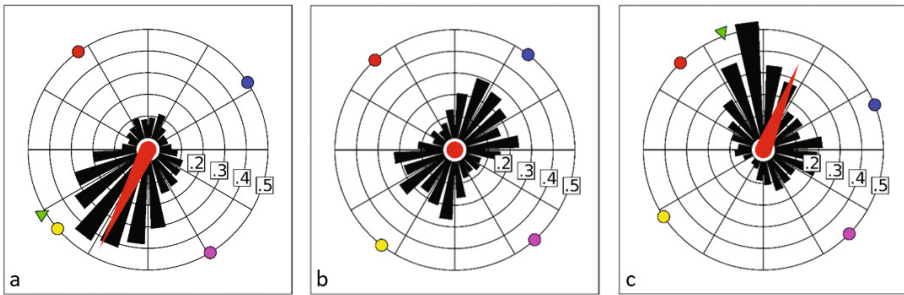


Fig. 10. Circular statistics of the bias direction for the control (**a.**), parallelogram (**b.**) and peak conditions (**c.**). The black columns show the approximate densities in expected cases per radian. The red needle shows the resultant vector, i.e. it points towards the circular mean while its length is a measure of concentration. The colored discs show the direction to the landmarks for each configuration. The green triangle indicates the bias direction predicted by the model. Note that for the parallelogram condition (**b.**) no bias is predicted, in agreement with the experimentally found resultant vector. (Color figure online)

bias predicted by the model (green triangle). Note that no bias is predicted in the parallelogram condition, Fig. 7b. The red needles show the circular mean of the distributions. The orientation of the distributions towards the predicted bias direction (green triangles) was tested against the null hypothesis of non-oriented distribution using the circular V -test (Batschelet 1981) and reveals significant deviations from the null hypothesis for the control ($V(954) = 0.389, p < 10^{-4}$) and peaked conditions ($V(754) = 0.291, p < 10^{-4}$). Since the model does not predict a bias for the parallelogram condition, we tested this condition with the most distant landmark direction as a predicted bias, but no significant effect was found ($V(761) = 0.021, n.s.$).

We conclude that place recognition from distant landmarks is based on a comparison of two components, (i) a referential place code containing veridical landmark distances and inter-landmark angles, and (ii) a visual working memory of the complete surroundings with distance-dependent resolution and systematic depth compression. A simple model of these components is able to quantitatively predict the statistical distribution of decisions made by human subjects. Effects of approach direction can be modeled by increasing the variances of the less seen landmarks.

The most surprising result of this study is the systematic bias found for the asymmetric landmark configurations. This bias can be modeled if we assume that the landmark distance instantaneously perceived during place recognition is hyperbolically compressed while the landmark distance represented in long-term memory is not. We think that this assumption is justified since the long-term place code (the landmark positions $\mathbf{l}_1, \dots, \mathbf{l}_4$) is a result of many encounters of the goal location arriving from all four directions. It is thus a consolidated memory taking into account multiple views and motion parallax during approach. In contrast, the perception during recognition is mostly instantaneous, with only limited access to the depth cues provided by motion parallax.

4 Conclusion

Place recognition is a simple, well-defined task in which a subject moves to a place (in reality or in virtual reality) and reports arrival by button hit. The main dependent variable is the observer position \mathbf{x} at button hit, to which simple bivariate statistics apply. Place recognition is interactive and continuous much like an adjustment task in classical psychophysics with the adjusted parameter being observer position. Button hit is triggered by a comparison operation involving memory content. In this respect, place recognition is more like a match-to-sample task in which the sample has to be remembered. It differs from match-to-sample in the continuity of the position parameter that allows gradual similarity. Also, the memory content may change during the experiment due to spatial updating processes accompanying the approach movement.

In this paper, we discussed two experiments on place recognition addressing the various stages of the appearance-based place recognition model of Fig. 1. The results are consistent with a simple model of spatial working memory (Eq. 4) making the following assumptions:

1. The landmarks are distinguished and identified (index i in model).
2. Both landmark distance and bearing are represented in an egocentric but geo-oriented reference frame.
3. Landmark distance in working memory is systematically biased according to hyperbolic distance compression (Eq. 1).
4. Landmarks outside the field of view are also represented at an updated position, but statistical error is larger than for actually perceived landmarks.

This last assumption, i.e. the increased error of representations of landmarks out of sight is not relevant for the results presented here, but has been used to model the effects of different approach directions by Lancier (2016).

Similar assumptions are made in most models of spatial working memory. For example, Loomis et al. (2013) assume that object knowledge is maintained and updated in a spatial working memory. While this is well in line with our results, it does not lend itself easily for quantitative predictions, as are sought in this study. The Byrne et al. (2007) model assumes a map-like representation in which walls or other objects are represented as activity in pixels of the map. The model nicely explains spatial updating but it is not obvious how to represent object identities. In contrast, object identities are easily accounted for in the Röhrich et al. (2014) model which is based on views of the environment and therefore automatically represents visual landmark properties. However, it lacks a mechanism for spatial updating which would have to be based on some sort of ego-motion dependent view transformations.

In summary, our results call for an improved model of spatial working memory, accommodating both object identities and spatial updating in a way allowing quantitative predictions.

References

- Basten, K., Mallot, H.A.: Simulated visual homing in desert ants natural environments: efficiency of skyline cues. *Biol. Cybern.* **102**, 413–425 (2010)
- Batschelet, E.: *Circular Statistics in Biology*. Academic Press, London (1981)
- Burgess, N., Hitch, G.: Computational models of working memory: putting long-term memory into context. *Trends Cogn. Sci.* **9**, 535–541 (2005)
- Byrne, P., Becker, S., Burgess, N.: Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychol. Rev.* **114**, 340–375 (2007)
- Cartwright, B.A., Collett, T.S.: How honey bees use landmarks to guide their return to a food source. *Nature* **295**, 560–564 (1982)
- Cheng, K., Huttenlocher, J., Newcombe, N.S.: 25 years of research on the use of geometry in spatial reorientation: a current theoretical perspective. *Psychon. Bullet. Rev.* **20**, 1033–1054 (2013)
- Dittmar, L., Stürzl, W., Baird, E., Boeddeker, N., Egelhaaf, M.: Goal seeking in honeybees: matching of optic flow snapshots? *J. Exper. Biol.* **213**, 2913–2923 (2010)
- Epstein, R., Kanwisher, N.: A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998)
- Epstein, R.A.: Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* **12**, 388–396 (2008)
- Franz, M.O., Schölkopf, B., Mallot, H.A., Bühlhoff, H.H.: Where did I take this snapshot? Scene-based homing by image matching. *Biol. Cybern.* **79**, 191–202 (1998)
- Gaussier, P., Revel, A., Banquet, J.P., Babeau, V.: From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. *Biol. Cybern.* **86**, 15–28 (2002)
- Gilinsky, A.S.: Perceived size and distance in visual space. *Psychol. Rev.* **58**, 460–482 (1951)
- Gillner, S., Mallot, H.A.: Navigation and acquisition of spatial knowledge in a virtual maze. *J. Cogn. Neurosci.* **10**, 445–463 (1998)

- Gillner, S., Weiß, A.M., Mallot, H.: Visual place recognition and homing in the absence of feature-based landmark information. *Cognition* **109**, 105–122 (2008)
- Graham, P., Cheng, K.: Ants use panoramic skyline as a visual cue during navigation. *Current Biol.* **19**, R935–R937 (2009)
- Halfmann, M.: Place Recognition and Navigation in Virtual Environments. PhD thesis, Faculty of Science, University of Tübingen, Germany (2016)
- Hamilton, D.A., Sutherland, R.J.: Blocking in human place learning: Evidence from virtual navigation. *Psychobiology* **27**, 453–461 (1999)
- Hartley, T., Maguire, E.A., Spiers, H.J., Burgess, N.: The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron* **37**, 877–888 (2003)
- Hollup, S.A., Molden, S., Donnett, J.G., Moser, M.-B., Moser, E.I.: Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *J. Neurosci.* **21**, 1635–1644 (2001)
- Hort, J., Laczó, J., Vyhnaček, M., Bojar, M., Bures, J., Vlček, K.: Spatial navigation deficit in amnesic mild cognitive impairment. *Proc. Nat. Acad. USA* **104**, 4042–4047 (2007)
- Janzen, G., van Turenout, M.: Selective neural representation of objects relevant for navigation. *Nat. Neurosci.* **7**, 673–677 (2004)
- Kuipers, B.: The spatial semantic hierarchy. *Artif. Intell.* **119**, 191–233 (2000)
- Lambrinos, D., Maris, M., Kobayashi, H., Labhart, T., Pfeifer, R., Wehner, R.: An autonomous agent navigating with a polarized light compass. *Adapt. Behav.* **6**, 131–161 (1997)
- Lancier, S.: Spatial Memories in Place Recognition. PhD thesis, Faculty of Science, University of Tübingen, Germany (2016)
- Loomis, J.M., Klatzky, R.L., Giudice, N.A.: Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language. In: Lacey, S., Lawson, R. (eds.) *Multisensory Imagery: Theory and Applications*, pp. 131–156. Springer, New York (2013). https://doi.org/10.1007/978-1-4614-5879-1_8
- Meilinger, T.: The network of reference frames theory: a synthesis of graphs and cognitive maps. In: Freksa, C., Newcombe, N.S., Gärdénfors, P., Wölfl, S. (eds.) *Spatial Cognition 2008. LNCS (LNAI)*, vol. 5248, pp. 344–360. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87601-4_25
- Möller, R., Vardy, A.: Local visual homing by matched-filter descent in image distances. *Biol. Cybern.* **95**, 413–430 (2006)
- Morris, R.G.M.: Spatial localization does not require the presence of local cues. *Learn. Motiv.* **12**, 239–260 (1981)
- O’Keefe, J.: The hippocampal cognitive map and navigational strategies. In: Paillard, J. (ed.) *Brain and Space*, pp. 273–295. Oxford University Press, Oxford (1991)
- O’Keefe, J., Nadel, L.: *The Hippocampus as a Cognitive Map*. Clarendon, Oxford, England (1978)
- Papi, F. (ed.): *Animal Homing*. Chapman and Hall, London (1992)
- Philbeck, J.W., Loomis, J.M.: Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 72–75 (1997)
- Pickup, L.C., Fitzgibbon, A.W., Glennerster, A.: Modelling human visual navigation using multiview scene reconstruction. *Biol. Cybern.* **107**, 449–464 (2013)
- Röhrich, W., Hardiess, G., Mallot, H.A.: View-based organization and interplay of spatial working and longterm memories. *PlosONE* **9**(11), e112793 (2014)

- Rolls, E.T., Treves, A., Robertson, R.G., Georges-François, P., Panzeri, S.: Information about spatial view in an ensemble of primate hippocampal cells. *J. Neurophysiol.* **79**, 1797–1813 (1998)
- Schindler, A., Bartels, A.: Parietal cortex codes for egocentric space beyond the field of view. *Current Biol.* **23**, 177–182 (2013)
- Schölkopf, B., Mallot, H.A.: View-based cognitive mapping and path planning. *Adapt. Behav.* **3**, 311–348 (1995)
- Siegel, A.W., White, S.H.: The development of spatial representations of large-scale environments. In: Reese, H.W. (ed.) *Advances in Child Development*, vol. 10, pp. 9–55. Academic Press, New York (1975)
- Sperling, G., Landy, M.S., Doshier, B.A., Perkins, M.E.: Kinetic depth effect and the identification of shape. *J. Exp. Psychol. Hum. Percept. Perform.* **15**(4), 826–840 (1989)
- Tinbergen, N., Kruyt, W.: Über die Orientierung des Bienenwolfes (*Philanthus triangulum* Fabr.) III. Die Bevorzugung bestimmter Wegmarken. *Zeitschrift für vergleichende Physiologie*, vol. 25, pp. 292–334 (1938)
- Tolman, E.C.: *Purposive behavior in animals and men*. The Century Company, New York (1932)
- Waller, D., Loomis, J.M., Golledge, R.G., Beall, A.C.: Place learning in humans: The role of distance and direction information. *Spatial Cogn. Comput.* **2**, 333–354 (2000)
- Wiener, J.M., Mallot, H.A.: ‘Fine-to-coarse’ route planning and navigation in regionalized environments. *Spatial Cogn. Comput.* **3**, 331–358 (2003)
- Wilson, M.A., McNaughton, B.L.: Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993)
- Zeil, J., Hoffmann, M.I., Chahl, J.S.: Catchment areas of panoramic snapshots in outdoor scenes. *J. Opt. Soc. Am. A*: **20**, 450–469 (2003)