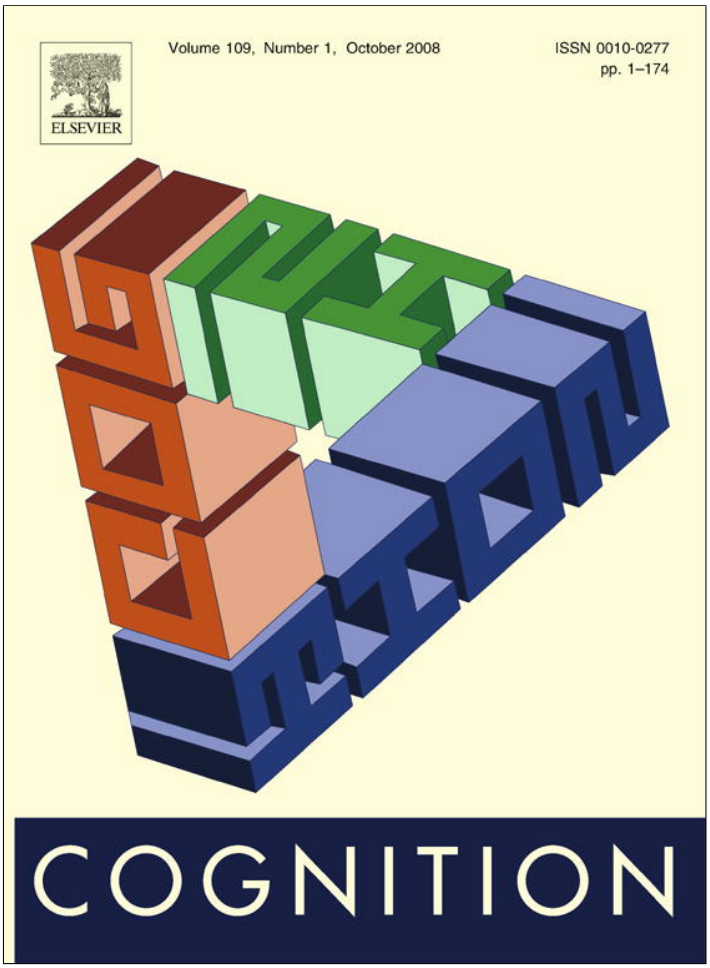
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Cognition 109 (2008) 105–122

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Visual homing in the absence of feature-based landmark information

Sabine Gillner, Anja M. Weiß, Hanspeter A. Mallot *

Department of Cognitive Neurobiology, Faculty of Biology, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany

ARTICLE INFO

Article history:
Received 17 March 2008
Revised 17 July 2008
Accepted 18 July 2008

Keywords:
Human navigation
Landmark
Snapshot memory

ABSTRACT

Despite that fact that landmarks play a prominent role in human navigation, experimental evidence on how landmarks are selected and defined by human navigators remains elusive. Indeed, the concept of a 'landmark' is itself not entirely clear. In everyday language, the term landmark refers to salient, distinguishable, and usually nameable objects, rendering the problem of landmark recognition a special case of the general object recognition problem. In contrast, in the insect and robot literature, this notion of landmarks is often replaced by the "local position information" [e.g., Trullier, O., Wiener, S., Berthoz, A., & Meyer, J.-A. (1997). Biologically based artificial navigation systems: Review and prospects. *Progress in Neurobiology, 51*, 483–544], referring to the entire set of sensor readings obtained at one location. This set may then serve as a characteristic of the particular location. Honey bees have been shown to base place recognition and homing on a snapshot-like memory of the place's visual environment, not on the distances to recognized objects [Cartwright, B., & Collett, T. (1983). Landmark learning in bees. Experiments and models. *Journal of Comparative Physiology A, 151*, 521–543]. A number of theoretical models of snapshot-based homing [e.g., Franz, M., Schölkopf, B., Mallot, H. A., Bülthoff, H. H. (1998). Where did I take that snapshot? Scene-based homing by image matching. *Biological Cybernetics, 79*, 191–202; Vardy, A., & Möller, R. (2005). Biologically plausible visual homing methods based on optical flow techniques. *Connection Science, 17*, 47–89] predict that the accuracy of snapshot-based homing should depend on image contrast. For rodent hippocampal place fields, models have been developed using additional image information such as three-dimensional depth and allocentric orientations (e.g., room axes) and are thus less sensitive to image contrast and noise [e.g. Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., et al. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences, 17*, 71–79]. Here, we study human visual homing in a virtual environment void of objects and readily detected image features. The environment was a circular room with a homogenous colour gradient covering the wall and uniform floor and ceiling. Subjects were able to approach remembered places. Accuracy decreased with colour gradient modulation and room size, in qualitative agreement with the snapshot model but not with other models of place recognition. We conclude that human memory for places can make use of a snapshot algorithm.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The recognition of places requires a memory of perceptions, visual or other, which characterize a particular location. Such memorized features of the environment are generally called cues or landmarks, even though the concept of a landmark remains vague and is used with different meanings within the interdisciplinary field of spatial cognition. Lynch (1960), a geographer, described landmarks as punctuate reference points which are usually physical objects and contrasted them to other geographical concepts such as nodes, edges, paths and districts. The

* Corresponding author. Tel.: +49 7071 297 8830; fax: +49 7071 29 2891.
E-mail address: hanspeter.mallot@uni-tuebingen.de (H.A. Mallot).
URL: http://www.uni-tuebingen.de/cog (H.A. Mallot).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.07.018

Author's personal copy

psychologists Siegel and White (1975) defined landmarks as unique configurations of perceptual events thus emphasizing the point that not only single objects but also configuration of objects could serve as a landmark. Clearly, objects or percepts do not become landmarks by themselves but by being remembered and recognized by a navigating agent. Siegel and White (1975) also suggested that landmark knowledge should be conscious. This idea is taken up in naming experiments where verbal recall is used to identify landmarks out of "non-landmark" objects. (e.g. Allen, Kirasic, Siegel, & Herman, 1979; Appleyard, 1969; Daniel, Tom, Manghi, & Denis, 2003). However, it is not clear whether those objects which are readily named are the same which are used as landmarks in a navigation task. Quite to the contrary, geometry cues of rooms or intersections, which generally are not reported verbally, have been shown to support localisation (Hermer & Spelke, 1994; Janzen, Hermann, Katz, & Schweizer, 2000; Stankiewicz & Kalia, 2007).

A general discussion of landmarks can be organized along three dimensions, i.e., (i) depth of landmark processing, (ii) landmark selection and (iii) navigational landmark usage. In this paper, experiments will be presented addressing the most basic level of landmark processing, i.e., visual snapshots lacking identifiable image features.

1.1. Landmarks

1.1.1. Depth of landmark processing

One dimension along which landmarks may be distinguished is the amount of sensory processing involved, ranging from raw sensor readings with no or minimal processing, to feature maps, recognized objects, or symbolic landmark names. All conceivable position cues characterizing a given place are contained in the complete set of sensory data available at that place, i.e., in the *local position information* (Trullier, Wiener, Berthoz, & Meyer, 1997). For the visual modality, local position information is composed of the raw array of colour and intensity values in the retinal images obtainable from a given view point (Gibson's "ambient visual array", Gibson, 1979). Individual retinal images correspond to a view point plus a viewing direction which is known as the "pose vector" in the robotics literature (see, for example, Hübner & Mallot, 2007). By means of image processing routines, more elaborated information can be extracted from the images, including depth maps, or visually recognized objects, in closer correspondence to the common sense usage of the term "landmark".

1.1.1.1. Snapshots. The use of barely processed image information for guidance has been demonstrated in a number of insect species, most notably digger wasps (Tinbergen & Kruyt, 1938), honey bees (Cartwright & Collett, 1983) and wood ants (Durier, Graham, & Collett, 2003). In one experiment, honey bees are trained to visit a feeder in the vicinity of three landmark objects. After training, the landmarks are moved such that two candidate search locations can be predicted, one where the snapshots (relative bearings of landmarks) match the training arrangement and one where the distances to the landmarks match. Bees were found to preferably search at the location where the snapshots match. Snapshot-based place recognition may also explain spatial learning of rodents in the Morris water maze as well as similar performances by humans in virtual reality mazes (e.g., Jacobs, Thomas, Laurance, & Nadel, 1998; Kallai et al., 2007). Further evidence for the view-based usage of landmarks in human subjects has been presented by Mallot and Gillner (2000).

Snapshots are associated to the location at which they are taken, not to the location of objects they may happen to depict. They thus conform to Siegel and White's (1975) view of landmarks as being memories of perceptions rather than memories of actual real world objects. Consider as an example the location called "Four Lakes View" (in German: Vier-Seen-Blick) in the valley of the river Rhein in Germany (Fig. 1). Looking from this place, the view of the river is interrupted by several small mountain ridges and therefore appears as four separate "lakes". Clearly, the landmark memory in this case refers to a view attached to a named place, not an object in the environment. Similar cases may be found for view axes or the visual alignment of multiple objects.

1.1.1.2. Depth signature. Places may also be characterized by an array of distance estimates taken from the current position to the surrounding obstacles or walls. Here, the application of an early vision operations, i.e., depth perception, results in slightly more abstract place descriptors that still lack identifiable landmark objects but exhibit an increased invariance e.g. to changes in illumination of the scene (Stürzl & Mallot, 2002). Egocentric depth signatures are the main cue in experiments on the so-called geometric module, where rats have been shown to use information on the rectangular layout of an experimental box for finding a hidden food pellet (Cheng, 1986). In doing this, the rats ignored odours as well as strong visual features such as coloured box walls or identifiable visual landmarks, and confused diagonally opposed corners ("rotation error"). Still, place finding was based on visual cues, i.e., the distances to the surrounding walls, presumably perceived by the elevation in the visual field of the edge between wall and floor as seen from the rat's point of view. Cheng (1986) suggested that depth (or 'geometry') is processed in a module separate from other visual information. Recently, this modularity assumption was challenged by Graham, Good, McGregor, and Pearce (2006) who showed that both depth and colour information is jointly used in rat navigation. In humans, initial claims that young children rely exclusively on the geometric module (Hermer & Spelke, 1994) have not been confirmed in experiments using different absolute room sizes (Learmonth & Nadel, 2002). However, adults engaging in a secondary, cognitively demanding task (Hermer-Vasquez & Spelke, 1999) also confuse diagonally opposing corners in a rectangular room, indicating their preference for geometric layout over visual surface markings as information source. Irrespective of the problem of modularity, the results suggest that depth estimates obtained from an array of egocentric directions can be used as landmark information.

Firing characteristics of neurons in the rodent hippocampal place cell system are also partially based on

Author's personal copy

Fig. 1. The "Four Lakes View" close to the town of Boppard, Germany. When looking from this place, four separate sections of the Rhein river are seen, limited by occluding mountain ranges. The named landmarks, i.e., the four lakes, describe the view visible from this place, not actual lakes in the landscape (Image from: http://de.wikipedia.org/, "Vierseenblick").

landmark information (for review, see McNaughton, Battaglia, Jensen, Moser, & Moser, 2006). Barry et al. (2006) and Barry and Burgess (2007) presented a model for the combination of landmark and egomotion information in place cell firing, based on an allocentric depth signature. Relative to an allocentric reference direction defined for the current environment, distances to the surrounding boundaries (e.g., walls of a room) are probed at a number of directions. The distance of the wall in a given allocentric direction is called a "boundary vector". As the rat turns, the probing directions are fixed with respect to the allocentric reference direction, i.e., the surrounding environment. As in the "geometric module", distances may be sensed by the retinal elevation of the image of the wall's lower edge. The model was shown to predict place cell firing fields across changes of the environment. In terms of visual processing, it requires local depth estimates plus recognition of an allocentric axis.

Depth signatures make most sense in indoor environments. Janzen et al. (2000) showed that the angle at which two hallways meet is included in the subjects' representation of space. Stankiewicz and Kalia (2007) distinguished structural landmarks such as hallway junctions from object landmarks and showed that both types of landmarks are used in navigational tasks. Interestingly, structural landmarks are more readily encoded that object landmarks. In the depth signature representation, hallway crossings show up as local figure of plus shapes whereas flat images on the walls will not be encoded. Depth signatures also relate to the notion of view-sheds or isovists, used in architecture to characterize indoor environments (Wiener et al., 2007).

1.1.1.3. Object landmarks. In everyday language, the term landmark refers to nameable, identified objects such as famous, conspicuous buildings, etc. One advantage of object recognition in navigation might lay in the fact that it can make use of small but characteristic regions of images such as distant objects that would not substantially influence snapshot matching (Stürzl & Zeil, 2007; Zeil, Hofmann, & Chahl, 2003). There are, however, a number of important differences between object recognition in general and landmark recognition in particular. First, object recognition should be view-invariant whereas landmarks are better defined in a view-dependent way, including information on the approach direction during landmark encoding. In other cases view invariance might be helpful to guide a navigation process. For example, it may be advantageous to understand that a house currently approached from the South, say, is the same house encountered previously from the North. Still, North and South views must be distinguished and perceiving them will generally lead to different navigational decisions. The situation is thus different from standard object recognition where recognition invariant of viewing direction is sought. Second, landmark recognition (as well as face recognition) is always an individual level recognition. Recognizing a landmark as "a tower" or "an arch", rather than as "the Eiffel Tower" or "the Brandenburg Gate", would be of little help in navigation although these superordinate level recognitions would be very important in object recognition. Finally, landmarks need not correspond to nameable objects at all. Rather, cues like the curvature of a street (Heft, 1979), geographical slant (Restat, Steck, Mochnatzki, & Mallot, 2004) or the angle at which two streets meet (Janzen et al., 2000; Stankiewicz & Kalia, 2007) make perfect landmarks but will hardly be reported verbally.

1.1.2. Landmark selection: Salience, reliability, and relevance

The success of landmark-based navigation depends critically on the sensible selection of landmarks from the environment. In a study with 6–12-year-old children, Cornell, Heth, and Broda (1989) showed that the criteria for landmark selection must be learned: Children have the tendency to remember interesting objects irrespective of their usefulness for navigation, including objects which are far away or too small. If the attention of the children was directed to objects which have also been chosen by adults, the children's navigation performance improved.

Three criteria can be identified for the selection of landmarks: (i)*salience*, (ii) *reliability or permanence*, and (iii) *relevance*. First, landmarks should be *salient*, i.e., they should be clearly distinguishable from their surroundings. This point was already stressed by Appleyard (1969), who

Author's personal copy

showed that local contrast of an object with respect to its immediate surroundings is an important criterion for reporting it as a landmark. This contrast may occur in any sensory dimension, e.g., size, colour or texture. Saliency is, however, not sufficient for sensible landmark selection, as was shown in the study by Cornell et al. (1989).

The notion of landmark *reliability or permanence* refers to the likelihood of finding the same landmark again when returning to a place after some time. For example, a car parked at an intersection does not make a good landmark even though it may be salient and also navigationally relevant. In the rat head-direction system, Zugaro et al. (2004) found a preference for the more distant one of two landmarks covering the same visual angle from the rat's point of view, where distance was judged by motion parallax. The authors suggest that the ecological advantage of this preference may lay in the fact that the more distant landmark covering the same visual angle is the larger one in absolute measures and will therefore be less likely to move away.

The *relevance* of a landmark or the place characterized by this landmark can be defined as the landmark's or place's importance in making navigational decisions. In experiments by Allen et al. (1979), subjects learned a route via a slide presentation and were then asked to name objects with potential landmark value. It turned out that objects at decision points are named more frequently than objects along the route. This result has been confirmed by Aginsky, Harris, Rensik, and Beusmans (1997) who showed that environmental changes occurring close to decision points are more readily detected than changes along a route. Faster recall rates for objects at decision points were also reported by Janzen (2006) who asked subjects to learn a route through a maze containing objects at decision points and non-decision points. This difference could also be demonstrated on the basis of neural activation (Janzen & van Turennout, 2004). In a learning phase, human subjects had to remember objects from one class (toys) placed at decision points or at neutral locations along the route. Contrasts of fMRI responses to toys vs. non-toys correlated with activity in the right fusiform gyrus whereas contrasts of objects at or off decision points correlated with activity in the left and right parahippocampal gyri. The study thus reveals a dissociation between object and landmark processing, where "landmarks" are defined by navigational relevance.

1.1.3. Usage of landmarks

In navigation, landmarks or local position information are used for the recognition of, the approach to, and the heading-on from places. The latter two of these usages correspond to the distinction of guidance and direction suggested by O'Keefe and Nadel (1978) and Trullier et al. (1997). For the mechanism of *guidance* an organism moves such that the currently visible image or some aspect of this image is equal to previously stored image information of a certain location or route. An important example of guidance is visual homing where the approach to a place (the "home") is guided by memorized visual information acquired at that place (the snapshot). In *direction* (or recognition triggered response), recognition of a place (or view) triggers the execution of a certain, memorized action, such as "go left at the purple house". Waller and Lippa (2007) trained human subjects to navigate a route composed of 19 binary (left/right) decisions by either associating a directional decision with a landmark ("direction" in the sense of O'Keefe & Nadel, 1978) or by choosing a way marked by a beacon. They found that beacon (or guidance) usage of landmarks allows better learning than associative landmark usage. However, beacon-based knowledge was found to be less enduring that associative knowledge. This may indicate that landmarks involved in the two navigational mechanisms are represented differently. In a general discussion of the notion of a landmark, this may indicate that there are different types of landmark knowledge corresponding to different behavioural contexts. As a further behavioural context, this may also hold for human verbal direction giving, where nameable object landmarks clearly seem to be preferred over non-nameable landmarks that may be equally useful navigationally (e.g., the structural landmarks of Stankiewicz & Kalia, 2007).

1.2. Snapshot memory in humans

While the existence of snapshot-based place memory is well established in insect research, little is known about the role of snapshot in human place finding and place recognition. The main reason for this is the problem of designing environments providing raw snapshot information but at the same time lacking identifiable landmark-objects. We addressed this problem by using an immersive virtual environment featuring a circular or square room with a smooth colour gradient covering the wall. As the gradient contained a full colour circle, no edges or discontinuities were present. In the vertical direction, wall colour was kept constant. Ceiling and floor were coloured in black and white, respectively (Fig. 2). We measured subjects' performance in homing to places within this environment depending on the contrast of the wall texture and the size and geometric layout of the room.

Algorithms working in the absence of identifiable landmark features rely on (i) local image contrast or (ii) local depth or distance to the objects. For each version of these algorithms, different predictions about homing performance can be made. In the original paper, Cartwright and Collett (1983) suggested that the raw images are transformed into edge images to gain some invariance with respect to illumination, changing colour of foliation, etc. In our circular environment, the only detectable edges are between the wall and the ceiling and between the wall and floor. For the circular room, edge based snapshots therefore predict a confusion of all places on a circle about the centre of the room, passing through the remembered place. Intensity-based mechanisms for snapshot matching using some sort of image correlation have been suggested e.g. by Franz, Schölkopf, Mallot, and Bülthoff (1998) and Vardy and Möller (2005). These algorithms are able to exploit the smooth texture information on the wall, but their performance will depend on texture contrast. An additional prediction made by all snapshot-based homing algorithms is that homing accuracy should decrease with decreasing

Author's personal copy

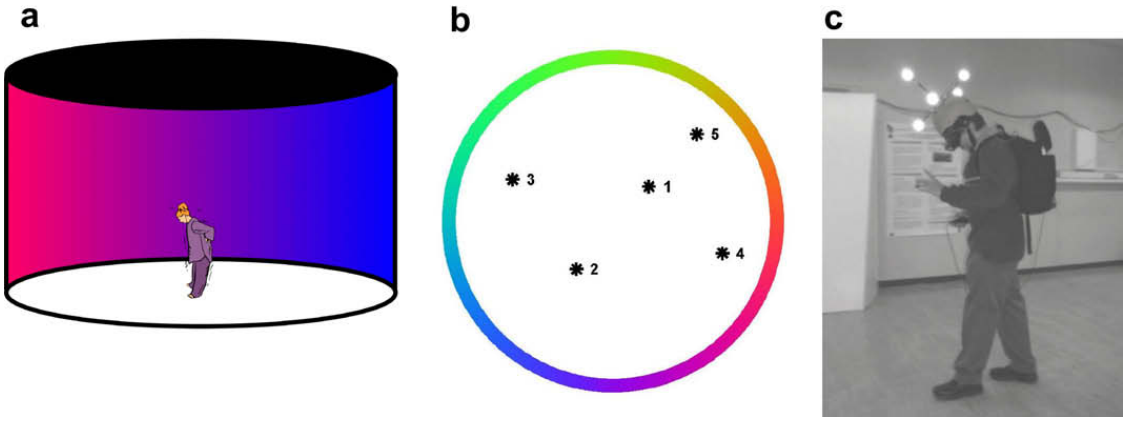


Fig. 2. Virtual environment. (a) Cylindrical room used in Experiment 1 (radius = 2.25 m, height = 4 m) with a smooth colour gradient covering the wall and uniform floor and ceiling. (b) Goal positions. While resting at one of five goal positions, subjects were allowed to view the room and memorize the location. Then they were teleported to another location and were asked to walk to the memorized place. The goal positions are fixed across all room sizes and shapes. (c) Subject in the simulated room wearing an HMD and a gauge object for the head tracker, consisting of five reflective balls mounted to a helmet. The balls appear luminous due to reflections from the photographic flashlight.

local image variation (Hübner & Mallot, 2007; Stürzl & Mallot, 2006). Clearly, when moving in the centre of a large featureless room, the snapshot will change less than for the same movement performed at the threshold of the entrance door. Therefore, the places in the centre will be less clearly characterized. Depth to the surrounding boundaries may be used in various ways. An egocentric depth model looking at the pattern of wall distances in egocentric coordinates predicts that positions on equidistant rings parallel to the wall should be confused. Note that this model is equivalent to the edge-based snapshot model of Cartwright and Collett (1983), at least for our circular environments; it is also equivalent to the depth signature approach of Stürzl and Mallot (2002). The relation of intensity- and edge-based snapshot models has recently been analysed in detail by Stürzl, Cheung, Chen, and Zeil (2008). Depth information might also be combined with image contrast, e.g. by assuming that a place is remembered by the colour of the closest wall segment, together with the distance from that wall segment (closest wall segment model). Finally, in the boundary vector cell model, wall distance is measured with respect to some allocentric room axis, such as the magenta-to-green axis in our virtual environment. All depth-based models predict that decreasing image contrast should not affect positional uncertainty with respect to the distance from the wall (radial position), since depth measurements are not affected by image contrast. Vice versa, positional uncertainty with respect to the tangential position (parallel to closest wall section) should depend on image contrast in the closest wall segment and boundary vector models.

1.3. Predictions

In this section, we will lay out the predictions for place recognition generated by the various models in a semi-quantitative way. The following models will be considered: (i) the image difference model (snapshot), (ii) the egocentric depth model (wall distances in egocentric directions), (iii) the closest wall segment model (colour and distance of closest wall segment) and (iv) the boundary vector cell model (allocentric room axis and wall distances in allocentric directions; Barry & Burgess, 2007; Barry et al., 2006). Since the only model in qualitative agreement with our data turns out to be the raw snapshot model, we will present more quantitative predictions from this model later in the paper.

1.3.1. Image based models

All models discussed are image based in the sense that they consist of (i) a place code or data vector which is close to the raw image, and (ii) a comparison or similarity function specifying the similarity between the codes of two places. Throughout this discussion, we assume that the arena is a circle with radius 1. We denote places and their coordinates by $g = (g_1,g_2)$ (goal) and $p = (p_1,p_2)$ (current position), and place codes by $c(g)$ and $c(p)$. The similarity function will be given by

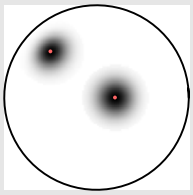
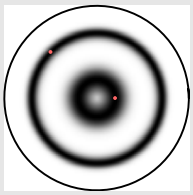
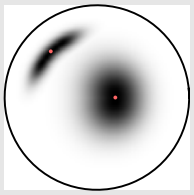
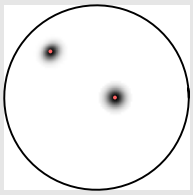
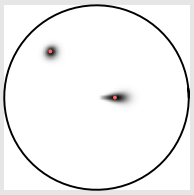
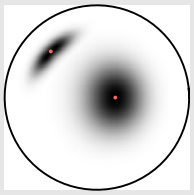
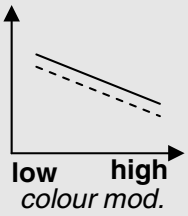
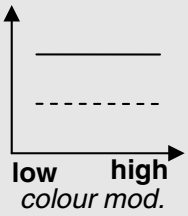
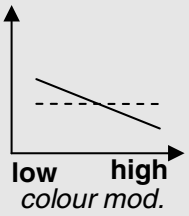
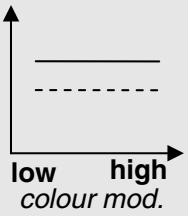
$$S(g,p) := \exp\left\{-\frac{1}{2}\sum_i \frac{||c_i(g) - c_i(p)||^2}{\sigma_i(g)^2}\right\}, \quad (1)$$

where $\sigma_i(g)$ is the standard deviation of the i-th component of the code vector representing g. The code functions will be specified for each model below. Table 1 shows a semi-quantitative account of four possible models by plotting the similarity functions for two goal points marked in red. The grey areas model the expected confusion areas, i.e., for each place, the shading corresponds to the likelihood of confusing this place with the respective goal. The diameters of the confusion area measured in radial (centre to goal) and tangential directions, averaged over all goal position for a given colour modulation level, are sketched schematically in the bottom row of Table 1.

In the *image difference model*, place-codes are simply the images obtained from each location, $c(\psi) = I(T_p(\psi))$, where $T_p(\psi) =: \varphi$ denotes the bearing obtained at the room centre of a wall point appearing at direction ψ when viewed from p. (Note that we consider here only the image intensities on a horizontal ring of pixels. A two-dimensional extension will be presented in the more quantitative simulations of

Author's personal copy

Table 1
Semi-quantitative predictions of confusion areas for two places (red dots) and four models of place recognition, using different amounts of colour modulation

Colour modulation	Image difference	Wall distance	Closest wall segment	Boundary vector model
Low				
High				
Expected var_t (—), var_r (- -)				

For explanation see text. Top two rows: predicted confusion areas for the various models at high and low image contrast (colour modulation). Bottom row: expected width of confusion area in the radial (var_r) and tangential directions (var_t). (For calculation of radial and tangential width from data, see below, results of experiment 2.)

Figs. 11 and 12 and, 16 below.) For a circular room with radius 1 and the centre at coordinates (0,0), simple trigonometry yields $\varphi = \psi + \sin^{-1}(p_2\cos\psi - p_1\sin\psi)$. Each code vector has the dimensionality of the number of image pixels. In the simulation, we used a numerical integration routine working on continuous image functions. The standard variances σ_i are assumed to be constant over all pixels (code dimensions). Simulations of $S(g,p)$ are shown for $\sigma = 0.07$ (high contrast) and $\sigma = 0.12$ (low contrast). Confusion areas are circular in all cases, they increase in size for lower colour modulation and they decrease in size towards the margin of the environment. Radial and tangential width both decrease in parallel for lower colour gradient modulation.

In the *wall distance model*, places are coded solely by their distance to the wall, $c(g) = 1 - (g_1^2 + g_2^2)^{0.5}$, i.e., by a single number with standard deviation assumed to increase with wall distance, $\sigma = 0.1 + 0.4c$. This assumption is equivalent to a logarithmic compression of perceived depth and constant noise (see Barry et al., 2006). Note that in circular environments, it does not matter whether distance is measured only in the direction to the closest wall or in all directions. In any case, confusion areas are circular rings concentric to the room margin. Clearly, confusion areas do not depend on colour modulation in this model.

For the *closest wall segment model*, places are encoded by two numbers, the distance to the wall $c_1(g) = 1 - (g_1^2 + g_2^2)^{0.5}$ as before, and the wall colour obtained at the closest wall segment. It is encoded by its angular position in the room coordinate system, $c_2(p) = \arg(p)$, where the arg-function denotes the quadrant-specific inverse tangent. Standard deviation of wall distance is modelled as before, $\sigma_1 = 0.1 + 0.4c_1$. For the standard deviation of the angular position of the closest wall segment, we assume a dependence on wall distance and on colour modulation. In Table 1, we used $\sigma_2 = \sigma_1/0.5$ (low colour modulation) and $\sigma_2 = \sigma_1/0.8$ (high colour modulation). In this model, there is a substantial dependence of confusion areas on colour modulation. In addition, the model predicts that radial and tangential width vary with colour modulation in different ways.

In the *boundary vector model* (Barry et al., 2006) places are coded by probing the distance to the wall in a number of fixed allocentric directions. An allocentric reference direction must be determined by landmark cues, in our case by the colour distribution on the wall. E.g., we may pick the magenta-to-green axis of the room as a reference direction. In this case, the place codes will also depend on colour modulation, since the recognition of the reference direction will be fuzzy if colour modulation is low. We modelled this case by encoding places by wall distances in eight regularly spaced allocentric directions of space. Each distance estimate is affected by noise with distance-dependent standard deviations as before. Since the observer cannot be sure about the reference direction, a fan of seven directions centred about the actual allocentric reference direction was calculated spanning a width of ±5 degrees in the high colour modulation case and ±12 degrees in the low colour modulation case. For each point p, the similarity function over the eight-dimensional code vector is calculated seven times, sampling the fan of possible reference directions. The overall place code difference is taken to be the minimum of these seven place similarity calculations. Thus, if the reference vector is uncertain, places on a

Author's personal copy

concentric circle in the arena may be confused since the boundary vectors may be rotated so as to fit to the actual wall distances also at neighbouring places. Results presented in Table 1 indicate that in the boundary vector model, dependence on colour modulation is low. Eccentricity of confusion areas depends strongly on wall distance.

In summary, we can make the following predictions:

P1 Both snapshot and depth-based models predict that confusion areas should increase in large rooms. For the depth-based models, this feature results from the variance of the individual depth measurement which is assumed to increase with absolute depth.

P2 Intensity-based snapshot models (Franz et al., 1998) predict a positive correlation between homing accuracy and contrast of the environment. Confusion areas are roughly circular over all colour modulation levels and a wide range of goal locations, i.e., radial and tangential variance are affected by contrast in the same way.

P3 Egocentric wall distance models and edge-based snapshot models (Cartwright & Collett, 1983) predict that homing should be better in rectangular rooms. Dependence on image contrast occurs only for the tangential width of the confusion area, not for its radial width.

P4 The boundary vector cell model (Barry et al., 2006), as all depth-based models, predicts that radial variance of the confusion area should not depend on colour modulation, at least if depth is judged from the upper and lower margins of the wall which are always at high contrast. The contrast dependence of tangential variance is only marginal.

We will show later that the intensity-based snapshot model is the only one consistent with our data. We will therefore present more quantitative predictions from that model below.

2. Methods

We performed navigation experiments in virtual reality. A total of 18 healthy adults (9 females and 9 males) participated in two experiments. This research follows the tenets of the Declaration of Helsinki. All participants were naive concerning the goal of our study; they were paid 8 €/h.

Subjects were asked to find certain locations within a simulated, or virtual, indoor environment built on a circular or square ground plane (Fig. 2). Six subjects, three male and three female, participated in the first experiment, while 12 participated in the second experiment.

2.1. Hardware

Subjects were allowed to walk and move freely inside a 6 m × 8 m walking arena, while wearing a head mounted display (HMD: Sony LCD Display: LDI – D100BE, Field of view: 33° × 22.7°). We recorded the movement of the subjects with an infrared-based 6-degrees-of-freedom movement tracker (A.R.T. GmbH, Weilheim, Germany) with an accuracy error of less then 0.1 mm. Subjects were wearing a reflective target, i.e., an arrangement of five balls covered with reflective foil (Fig. 2c). The simulated room was displayed in the HMD. Position and heading data from the tracker were transmitted to the graphic computer by wireless Ethernet with a temporal resolution of 30 Hz. These data were used to change the view point in the virtual reality according to the movement of the subject in the real world. The graphics computer was a laptop PC with NVidia Gforce2Go graphics card carried in a backpack by the subject, the refresh rate for the computer graphic was 30 Hz. The same computer was also used for data recording. Thus, subjects could move freely without cable connection to a base station.

2.2. Virtual environments

During the experiments, participants were asked to find certain locations within a cylindrical or a square simulated room. The experiments have been programmed by using C++ and OpenPerformer as graphic library, using Suse Linux 7.2 as operating system.

In Experiment 1, a circular room was used with a fixed diameter of 4.5 m and a floor-to-ceiling height of 2.5 m. Within the room, we use a coordinate system originating at the room centre with the x-axis pointing in the $\varphi = 0$ direction of the colour cycle. In this coordinate system, five goal positions were defined as specified in Table 2. In Experiment 2, additional circular rooms and square rooms were used. The circular rooms had diameters of 4.5 m, 9 m, 18 m and 27 m, respectively. The square rooms had side-lengths of 5 m, 9 m, 18 m, and 27 m. For mapping the colour gradient to the straight walls of the square room, the cycle was cut into four equal sections which were scaled and unfolded. The goal positions remained at the same coordinates for all room shapes and sizes.

The experimenter prevented participants from leaving the actual walking arena of 5.2 m by 6 m, corresponding to the tracked area of the physical room. Inside the simulated room, no objects or landmarks have been located. As the sole landmark cue, the walls were covered with a continuous colour gradient spanning the entire colour cycle (Figs. 2 and 3). The gradient was generated using three sinusoidal functions, one for the R, G, and B colour channels, respectively, with a phase shift of 120° between the channels:

$$f_i(\phi) = \frac{1 + c\sin(2\pi(\phi + i/3))}{2} \quad (2)$$

Table 2
Coordinates (x,y) of the goals as depicted in Fig. 2b with respect to the room centre at (0,0)

Goal	x [m]	y [m]	D [m]	Chance level [m]		
				Small circular room (r = 2.25 m)	Small square room (5 m × 5 m)	Larger rooms
1	0.51	0.49	1.5	1.61	2.00	2.22
2	−0.52	−0.68	1.4	1.66	2.04	2.26
3	−1.43	0.60	0.7	2.02	2.33	2.53
4	1.57	−0.45	0.6	2.07	2.37	2.57
5	1.20	1.25	0.5	2.14	2.43	2.61

By D, we denote the distance of each goal to the wall in the small cylindrical room. Also shown is the expected homing error, i.e., the chance level as calculated from Eq. (4); for further information see text.

Author's personal copy

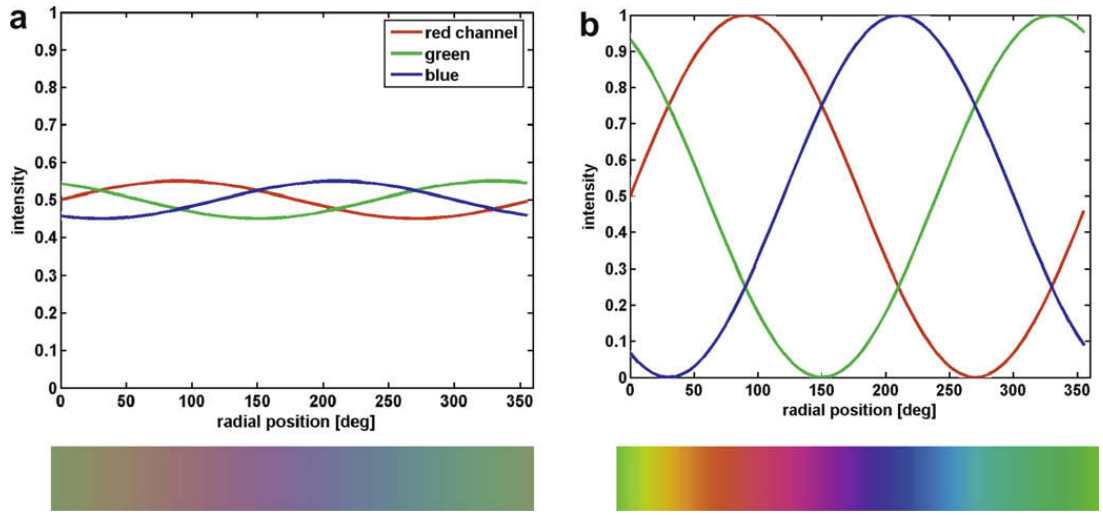


Fig. 3. Wall texture of the simulated rooms. Homogeneous colour circles of varying color modulation have been generated from sinusoidal intensity variations in the three colour channels (upper parts and Eq. (2)). The resulting texture is shown below. (a) Colour modulation (c) 10%, (b) colour modulation 100%.

where i takes the values 0, 1, and 2 for the red, green, and blue channels. We chose sinusoidal functions for the computation of the colour gradient since other functions resulted in marked discontinuities in the gradient due to color resolution limitations of the HMD, in particular at lower contrasts.

The variable φ in Eq. (1) ranges from 0 to 1 and denotes the angle of the colour cycle where 1 corresponds to 360°. Colour contrast was varied by adjusting the variable c controlling the amplitude of the underlying sinusoidal functions. In Experiment 1, we used the values $c = \{0.1, 0.2, 0.3, 0.4, 0.7, 1.0\}$. In Experiment 2, c was fixed to 1.

It should be noted that c is not a formal measure of colour contrast as defined in colourimetry (see, for example, D'Zmura & Singer, 1999), nor is the colour cycle equiluminant. For the purpose of the experiment, it was important to make sure that no localized features such as discontinuities in the colour gradient could be perceived. Put differently, the discriminability of two colours separated by an angle $\Delta\varphi$ should be equal for all absolute locations on the colour cycle. Therefore, discrimination thresholds were measured in a control experiment for three subjects. In a two-alternative forced choice task, subjects were presented with two colours from the sinusoidal gradient, separated by a distance $\Delta\varphi$. Using a standard staircase technique, discrimination thresholds for $\Delta\varphi$ were determined for a total of 15 equidistant points on the colour circle. Discrimination thresholds were shown to be roughly equal over the entire gradient, ensuring that no colour discontinuities are present that could be used as localized spatial cues (Fig. 4). The colour discrimination thresholds at different angular positions along the colour gradient differ only slightly, as revealed by a one way ANOVA ($F = 1{,}957$, $df = 14$, $p = .06$). A closer look reveals that only the thresholds at positions at $\varphi = 0.266$ and $\varphi = 0.533$ shows a slightly significant difference ($p < .1$, multiple comparison procedure after Milliken & Johnson, 1992) which was considered negligible.

2.3. Procedure

Each experimental run consisted of 20 homing trials and an initial training trial. The homing trials are carried out in two phases.

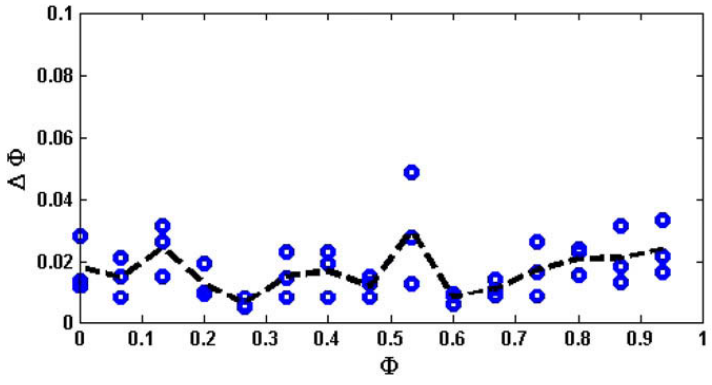


Fig. 4. Result of the colour discrimination experiment. φ denotes the position variable from Eq. (2). The thresholds $\Delta\varphi$ vary between 0.01 and 0.04, corresponding to angles of 3.6° to 14.4° (Hauser, 2005).

Author's personal copy

2.3.1. Goal presentation phase

A subject was standing in the walking arena at the start point or the end point reached in the preceding trial. The view-point was then moved to one of the five goal locations (Fig. 2b), while the subject was resting at his or her initial position ("teleportation"). Subjects had been instructed to memorize this location as accurately as possible. At the goal position, heading changes, but not translation movements, could be performed in closed-loop, i.e., upon turning the head, the appropriate parts of the visual field came into view. Thus, subjects were able to explore the entire "visual array" of the goal location. Memorization time was not restricted and ended when the subject pressed a button. The view-point was then moved back to the initial position.

2.3.2. Locomotion phase

After the goal presentation phase, virtual reality was operated in closed loop. Subjects were instructed to walk as accurately as possible to the target place encountered during the goal presentation phase. Again, there was no time limit for this. Subjects signalled the (perceived) arrival at the target position by pressing a mouse button. No feedback was given as to the correctness of this decision and the next trial was started from the chosen end point.

The initial trial was a training trial starting at a random position and leading to goal positions 1 or 2. The general procedure was as for the experimental trials except that the subjects received feedback after finishing the locomotion phase by presenting a button at the actual goal position. The data from the initial trial were excluded from further analysis.

From the five target locations, a sequence of 20 homing trials was generated including all possible target transitions in a closed trail (an Euler path of the completely connected 5-graph). Thus, every goal position was visited four times, once from each of the other goal position. The sequences used appear in Table 3; sequence 2 is a shifted version of sequence 1.

2.4. Design

2.4.1. Experiment 1

Six subjects performed one experimental run for each of six colour modulation conditions. The sequences in which the colour modulation conditions were performed were $c_1,c_2,\ldots,c_6$ for subject one, $c_2,c_3,\ldots c_6$, c_1 for subject two and so forth to $c_6,c_1,\ldots,c_5$ for subject six. Data were collapsed over colour modulation condition, eliminating possible learning effects. Each subject performed an experiment with one of two possible sequences for the homing trials (cf. Table 3). The factor sequence did not lead to a significant effect on homing accuracy (non-parametric Friedman test, $df = 1$, $\chi^2 = 0.02$, $p = .9$). The data from the two sequence conditions will therefore be pooled in later analysis.

2.4.2. Experiment 2

Twelve subjects participated in this experiment. They were randomly divided into two groups of six, balanced for gender, one group performing the square room condition and one performing the circular room condition. In each of these two shape conditions, four size conditions were tested, each with one experimental run per subject. Altogether, each subject performed four experimental runs, one in each size condition. As in the colour modulation experiment, subjects started with different room sizes, eliminating possible learning effects.

2.5. Data analysis

Complete head-trajectories of the subjects composed of 6 degree of freedom pose data (3D position and 3D orientation) were recorded at a sampling rate of 1 Hz. From these, head direction data could be obtained. From the endpoints of the trajectories, we generated scatter-plots with variance ellipses centred on the mean endpoint coordinates. Mean homing error was calculated by calculating the median of the Euclidian distances between each endpoint and the goal position. Since the distribution of these errors is not normal, we used non-parametric tests to analyse the data.

2.5.1. Image difference model

The empirical results will be compared with predictions from the root mean square model of image similarity as used by Stürzl and Mallot (2006), Zeil et al. (2003), Stürzl and Zeil (2007), and Stürzl et al. (2008). In this model, positions are encoded and stored as panoramic snapshots visible from each position. Deviations from the goal position are detected and quantified by the associated image differences. As a measure of image difference, we use the square root of the sum of the squared differences per pixel and colour channel. Since these differences depend strongly on the angular orientation of the snapshots, we orient all images towards one standard direction given in world coordinates ("North"). This amounts to the assumption that subjects are able to compensate for unintended image rotations when judging the image difference. For the computation of the theoretical results we took 360°-roundshots of the experimental VR-environment (Fig. 5) with a resolution of 800 × 300 pixels; the orientation of the camera was equal in all these images. These correspond to the places' visual arrays rather than to the actual views visible in the HMD, which had limited field of view. For every environment we took images at the positions of each goal

Table 3
Trial sequences (sequence I and sequence II) of goal positions for one experimental run

	Place sequence																				
I	2	4	1	5	3	1	2	5	4	3	2	3	4	5	2	1	3	5	1	4	2
II	1	2	5	4	3	2	3	4	5	2	1	3	5	1	4	2	4	1	5	3	1

The two versions were used in two "sequence conditions". The gray column marked the initial training trial which was excluded for further analysis.

Author's personal copy

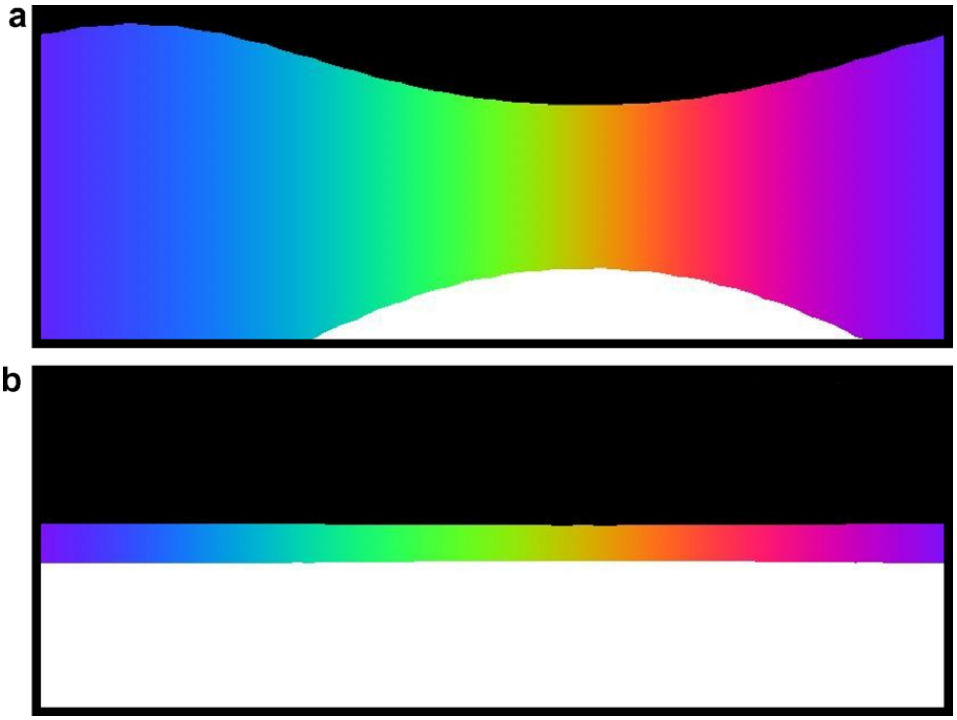


Fig. 5. Sample images used for the image difference calculations. (a) 360°-snapshot of the 100% colour modulation environment (S-Size). (b) 360°-snapshot of the XL-environment. The portions of the image showing ground and ceiling ground are much larger than in the smaller environment.

position and at additional 1949 positions, equally distributed in the environment. The separation of the positions p was 0.05 m in x- and in y-direction. The mean image difference d of each of theses images I_p to a certain goal image I_g was computed as

$$d(g,p) = \sqrt{\frac{1}{3 \times 800 \times 300} \sum_{c=1}^{3} \sum_{i=1}^{800} \sum_{j=1}^{300} (I_g(c,i,j) - I_p(c,i,j))^2}, \quad (3)$$

where c denotes the colour channel and i, j number the pixels in x- and y-direction, respectively.

From Eq. (3) we obtain an image difference d for each point in the environment p to each goal g. In the simulations, we assume that subjects recognize the goal position as soon as the image difference drops below a threshold d_o which was set to 0.08. From all points satisfying $d(g,p) < d_o$, we calculated a variance ellipse that can be compared with the subjects' performance. The threshold d_o was selected to match the overall size of the empirical variance ellipses.

2.5.2. Chance level

We computed the chance level for the homing error by assuming that no visual information whatsoever is used. If the subjects were wandering around without any clue, we expect them to indicate arrival at the goal g at random positions p. In this situation, the expected distance error $e(g)$ for each goal position can be calculated as the integral

$$e(g) = \frac{1}{|A|} \int \int_A \|g - p\| \mathrm{d}p, \quad (4)$$

where A denotes the walking arena, limited either by a virtual wall (Experiment 1 and small room conditions of Experiment 2) or the limits of the tracked area in the physical room (large room conditions of Experiment 2). $|A|$ the area of the walking arena, g and p are considered as two-dimensional vectors and the double bars indicate the Euclidian norm. The expected mean distance errors for each goal location are shown in Table 2. The expected homing error is larger for goal locations closer to the circular wall and varies in the range of 1.61–2.14 m.

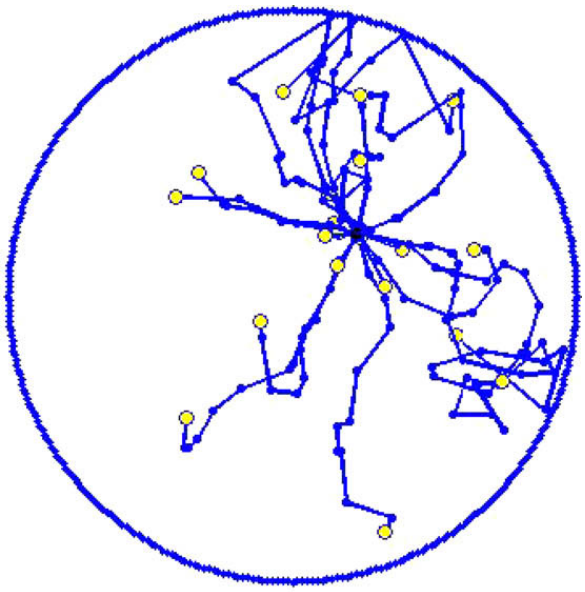


Fig. 6. Result of a time limited random walk, starting from goal No. 1. Step length as well as step number and angle between two adjacent steps were varied following a normal distribution. The resulting homing error is comparable to those obtained by the chance level, calculated from Eq. (4).

Author's personal copy

Chance levels and chance level dependence on room size and goal position were also confirmed by simulating a time limited random walk with normally distributed step length and turning angle (Fig. 6). For the random walk simulation, the step length, step number and angle between two adjacent steps very all varied according to a normal distribution with mean and variance estimated from the experimental data. The data are in overall agreement with the simpler chance level calculation given in Eq. (4) and Table 2.

3. Results

3.1. General

The experiments described in this paper measure the homing accuracy of human subjects in a computer graphics indoor environment without localized objects. In Experiment 1, we varied the modulation (or contrast) of the colour gradient covering the wall in the small room (diameter 4.5 m). In Experiment 2, we varied room size while colour modulation was kept constant at 100%. All accuracy data will be compared with predicted homing accuracies calculated from the squared image difference algorithm. One experimental session lasted between 45 min and one hour, one homing trial took on average 20.5 s in the first experiment and 43.6 s in the second experiment.

Sample trajectories from Experiment 1 (colour modulation) from one subject are shown in Fig. 7. Overall, subjects are able to home to memorized locations on a more or less straight path at least in the easier conditions (colour modulation of 100% or small room sizes). The viewing direction in the easier conditions is less variable than in the more difficult conditions (Fig. 8). The initial walking direction from the direction to the goal was calculated as a vector from the first to the fourth point of the trajectory (corresponding to the initial 4 s of the trajectory) The angular deviation of this vector from the goal vector pointing from the initial position to the goal is plotted in Fig. 9. It is around 45° in the easier experimental conditions and increases with room size but not with colour modulation.

3.2. Experiment 1: Colour modulation

Six subjects (3 male, 3 female) participated in this experiment. Colour modulation (variable c in Eq. (1)) was varied in six levels from 10% up to 100%. In a within-design, each subject performed 6 experimental runs, corresponding to the 6 colour modulation levels.

Fig. 10a and b shows the trajectory endpoints of all subjects in the 10% and 100% colour modulation conditions, together with their means and error ellipses. Also included in Fig. 10c and d are the variance ellipses predicted from the image-based theory (Eq. (3)) for the same colour modulation conditions. Image based theory predicts that the variance ellipses of homing locations should decrease for higher colour modulation levels.

This prediction was confirmed by the empirical data for the three goal locations closest to the wall. There was no reduction in homing error for the two central goals (Nos. 1 and 2), which are depicted in red and yellow in Fig. 10. The median of the homing error for all colour modulation levels is shown in Fig. 11a and b (experimental data) and Fig. 11c and d (prediction from image difference theory). The homing error is well below chance level (1.61–2.14 m, cf. Table 2) for all goals and all colour modulation levels. Homing error increases for lower colour modulation levels, the dependence of homing error on colour modulation is significant (non parametric Friedman test, $\chi^2 = 12.79$, $df = 5$, $p < .05$). There is also an effect of goal ($\chi^2 = 16.54$, $df = 4$, $p < .01$). If the data from the two central goals (1 and 2) are pooled and tested against the three peripheral goals (3, 4, and 5), a significant effect of goal eccentricity is found for the first two contrast levels ($c = 10\%$: $df = 1$, $\chi^2 = 4.08$, $p < .05$; $c = 20\%$: $df = 1$, $\chi^2 = 10.08$, $p < .01$). Smaller homing errors for the central

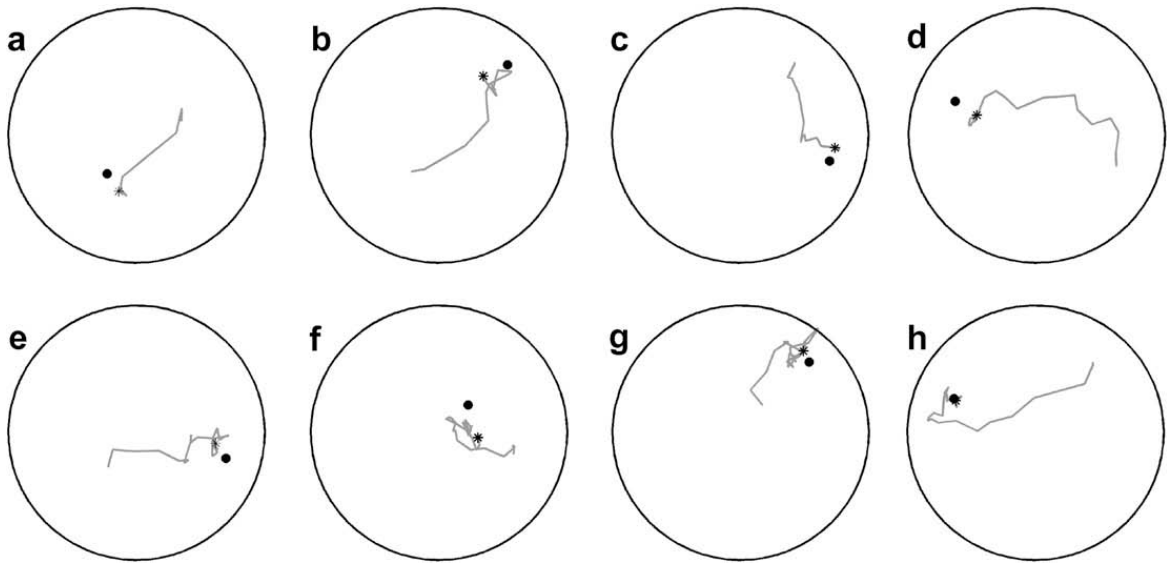


Fig. 7. Sample trajectories for one subject. (a–d) Four trials in the 100%-colour modulation environment and (e–h) in the 10%-colour modulation environment. The goals are marked by a black circle, the endpoint of a trajectory is marked by a star (*).

Author's personal copy

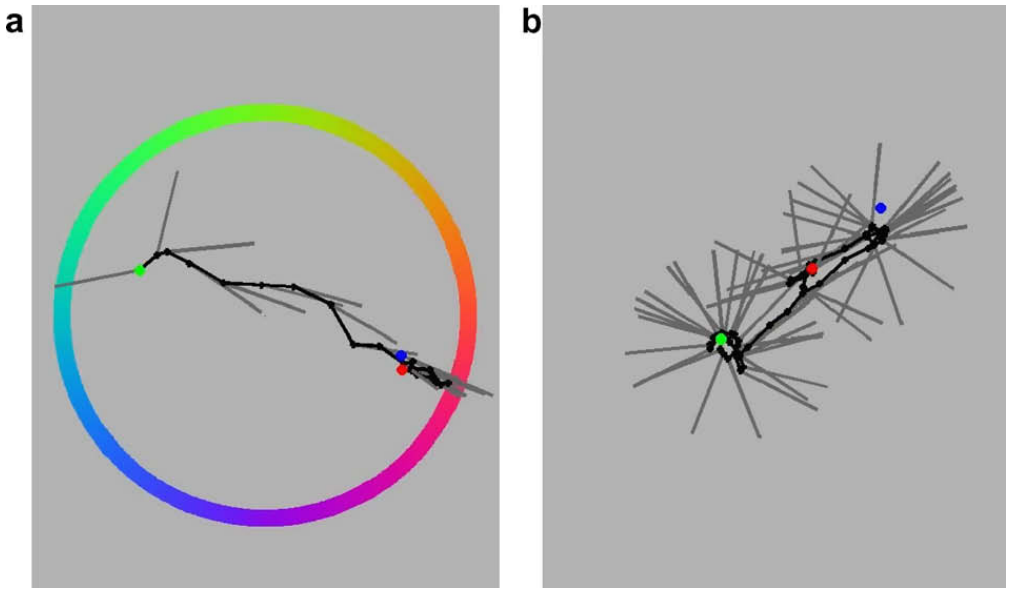


Fig. 8. Sample trajectories and head directions for one subject in the second experiment (different room sizes). Green dot: starting point, blue dot: goal, red dot: endpoint of trajectory. (a) Small circular room (S-environment). The colour circle corresponds to the position of the cylindrical wall, the grey rectangle marks the size of the tracked walking arena. (b) Large circular room (XL-environment), the cylindrical wall is outside the tracked arena.

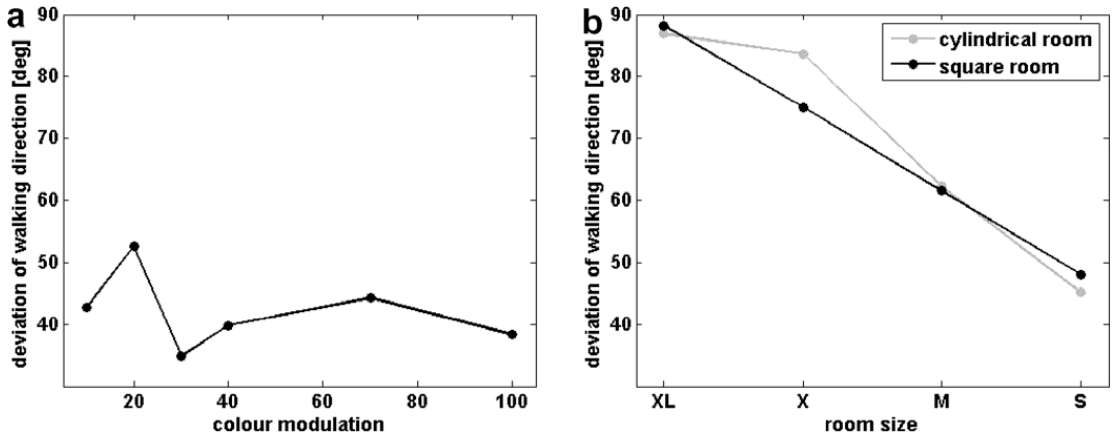


Fig. 9. Initial walking direction of homing trajectory. Plotted is the deviation of the initial walking direction from the bearing of the goal. (a) In the small rooms (expt. 1), the deviation is around 40–45 deg and does not depend on colour modulation. (b) In experiment 2, heading deviation is smaller in smaller rooms (Friedman test $df = 3$, $F = 51.29$, $p < .0045$) but does not depend on room shape.

as compared to the peripheral goals are predicted both by the theoretical chance level calculations (Table 2) and by the image comparison algorithm (Fig. 11d). However, chance level alone cannot explain the data since (i) the overall performance is well above chance and (ii) the effect of colour modulation is not predicted.

Fig. 12 shows the radial and tangential width of the confusion area as a function of colour modulation. Radial and tangential width was determined in the following way: For each data set (one goal location, one colour modulation level, 6 subjects, 4 trials, i.e., scatter plot of 24 points), the mean position or centre of gravity was calculated by separately averaging the x and y coordinate values. The direction from the room centre to this centre of gravity was called the radial axis, the orthogonal axis was called tangential. Next, each data point was projected to the radial and tangential axes, obtaining radial and tangential coordinates for each point. The standard deviation of the radial (tangential) values was then taken as a measure of the radial (tangential) width of the scatter field. Finally, the width values were averaged over all five goal locations from one colour modulation level. As colour modulation increases, both widths decrease roughly in parallel, as is revealed by a two way ANOVA (colour modulation $\times$ radial/tangential width) showing a significant main effect for colour modulation ($df = 5$, $\chi^2 = 3.93$, $p = .005$), a marginally significant main effect for radial vs. tangential measurements ($df = 1$, $\chi^2 = 4.62$, $p = .04$), and a marginally significant interaction ($df = 5$, $\chi^2 = 2.56$, $p = .04$). The theoretical prediction from the image difference model (Fig. 10c and d) show that confusion areas should get smaller for higher colour modulation but at the same time keep their spherical shape. This is indeed what we see in Fig. 12, where radial and tangential diameters of the confusion areas decrease in parallel. In contrast, both the closest wall segment model and the

Author's personal copy

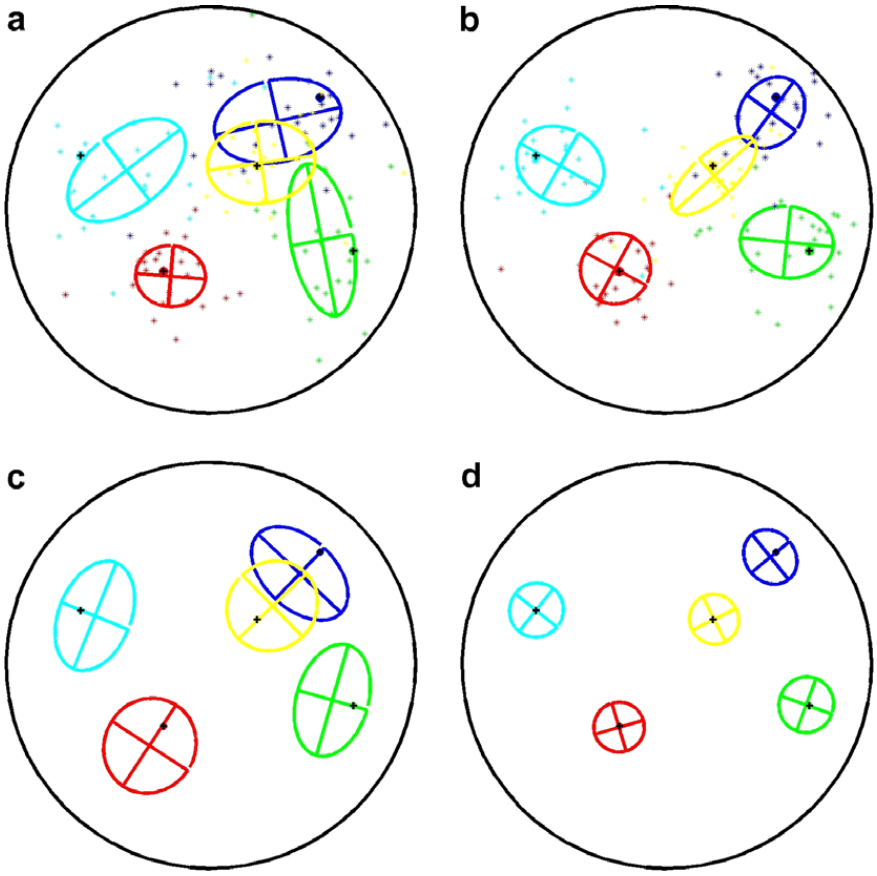


Fig. 10. Endpoints of the trajectories in Experiment 1. The black symbols mark the goal locations. (a and b) Empirical data (scatter-plot, means and error ellipses) (c and d) Simulation (only means and error ellipses are shown). Left column (a and c) colour modulation $c = 10\%$. Right column: (b and d) colour modulation $c = 100\%$. As predicted by the model, the variance ellipses of the empirical data are smaller with higher colour modulations.

boundary vector model predict that colour modulation should affect tangential, but not radial, width, yielding ellipsoidal or crescent-shaped confusion areas (cf. Table 1).

3.3. Experiment 2: Room size

For the second experiment, rooms of different sizes and shapes were used as described in Section 2 and in Fig. 13. Again we calculated the variance ellipses for both, the theoretical prediction and the empirical data, the results are depicted in Fig. 14. The theoretical model predicts a reduction of the homing accuracy in larger rooms. This result was confirmed by the experimental data. As can be seen in Fig. 15, the homing error is low for small rooms and increases to a level still well below chance level (2.3–2.5 m in the XL-environments, cf. Table 2) for the largest rooms. A Friedman test reveals a significant influence of room size ($df = 3$, $\chi^2 = 209.15$, $p << .01$) but no significant difference for the factor "shape" (circular vs. square room) ($df = 1$, $\chi^2 = 0.17$, $p = .68$). Therefore, we do not show data from the square room experiment graphically. In addition, a significant influence on the factor "goal" ($df = 4$, $\chi^2 = 20.45$, $p << 0.01$) was found.

4. Discussion

The experiments described in this paper use environments void of localized "cues" to show that visually guided place finding ("homing") can be based solely on smooth variations of colour and intensity in the visual snapshot. Homing accuracy decreases with colour contrast and with room size. Confusion areas around the goal locations are roughly circular or do at least not show preferred orientations with respect to the radial and tangential axes induced by the room geometry. Finally, a comparison of circular and square rooms did not reveal significant performance differences even though the corners of the square room may provide salient place cues. This finding indicates that localized cue information might even be neglected in some situations. We will now discuss these results in the light of the theoretical predictions summarized in Table 1.

4.1. Image-difference model

The results are consistent with the quantitative predictions of the image difference model. Homing accuracy decreases at lower colour contrasts and in larger rooms, at least for the more peripheral goal points. However, the exact shape of the curve showing the dependence of accuracy on contrast is only poorly matched by the prediction (Fig. 11). The reduction in colour modulation influenced homing accuracy only for modulation levels below 30%. This is not predicted by our simple image comparison model but is very plausible with respect to human perception. Detection and identification are known to be contrast dependent for small, sub-threshold contrasts, while

Author's personal copy

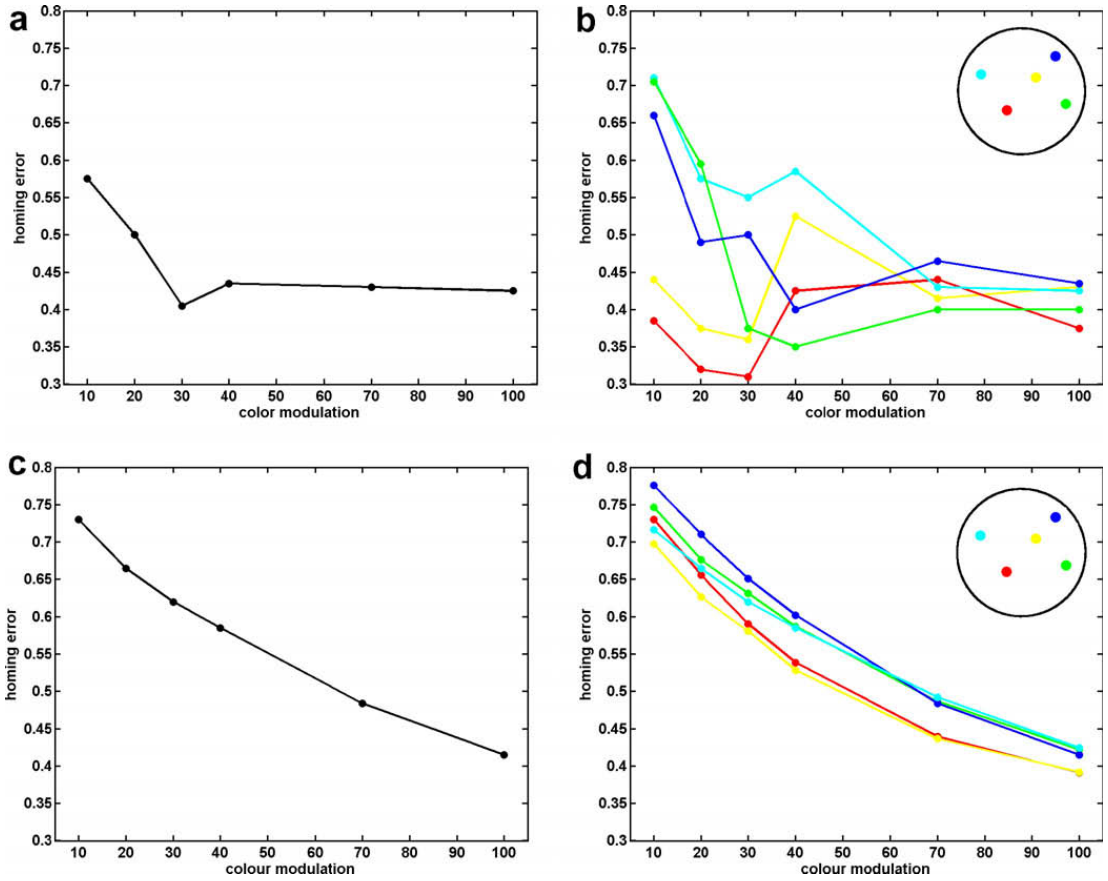


Fig. 11. Homing error as a function of colour modulation in Experiment 1. (a and b) Median of the experimental homing error in meter. (c and d) Simulation using the image difference model. The left column (a and c) shows simulation and data pooled over all goal locations. The right column (b and c) shows simulation and data separately by goal location. The goals are colour coded as shown in the inlay and in Fig. 10.

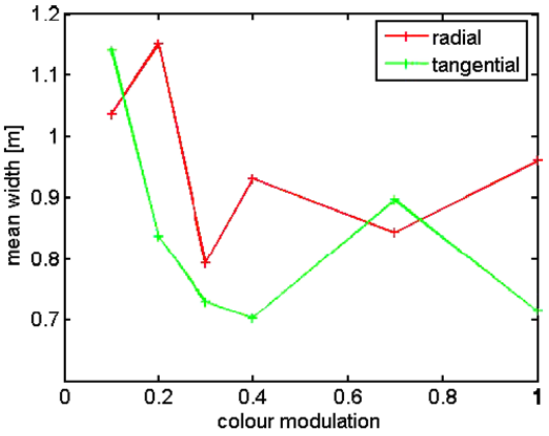


Fig. 12. Radial and tangential standard deviation of trajectory endpoints in Experiment 1. The standard deviations correspond to the diameters of the confusion area in tangential and radial directions. Both curves are roughly parallel, indicating that the confusion areas are approximately circular for all colour modulation values. This result should be compared to theoretical predications in Fig. 10c and d and Table 1 (bottom row).

increasing contrast beyond threshold will not make much difference (e.g. Campbell & Maffei, 1974). Better quantitative agreement of the behavioural data with the image difference model can probably be achieved by application of some nonlinear compression function on the image intensity data. However, the qualitative result, i.e., the monotonic decrease, seems to be the more relevant finding.

In terms of the individual goal points, the decrease in accuracy was found only for the peripheral locations, not for the central ones. One possible interpretation of this finding is that in the approach of the central points, depth-dependent strategies play a stronger role.

It could be argued that any navigational system that has to extract visual information might be sensitive to a reduction in contrast. It should be noted, however, that the colour modulation parameter *c* in out experiment does not affect the visibility of the edges between the wall and the floor and ceiling. Since we assume that these edges make the most salient cues to depth, depth-based models should not be affected by colour modulation.

A version of the image-difference model used in our simulations is the edge-based model of Cartwright and Collett (1983) where local edges are extracted prior to image comparison. Clearly, this model will be less sensitive to colour modulation change, but will be hardly able to extract any useful feature in our circular environments. In the square-room conditions, performance was not improved, even though the corners of the room do provide useful edge information. Thus, the additional edge information is either not used or does not lead to improvement.

The image difference model also predicts that the shape of the confusion areas should be roughly circular in all

Author's personal copy

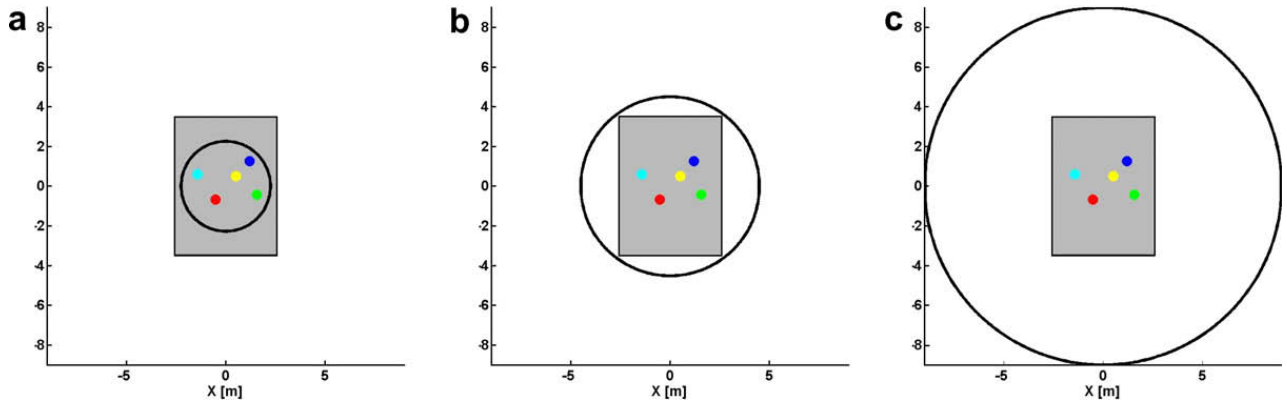


Fig. 13. Birds eye view of the S, M and L circular environments in the second experiment. The grey rectangle marks the size of the (physical) walking arena. The goal locations were equal in all conditions, while the diameter of the circular room has been varied (a) S-environment, 4.5 m, (b) M-environment, 9 m, (c) L-environment, 18 m. There was also an XL-environment (27 m), which is not shown here. In addition, we used four square environments of roughly equal size (see "Methods" Section).

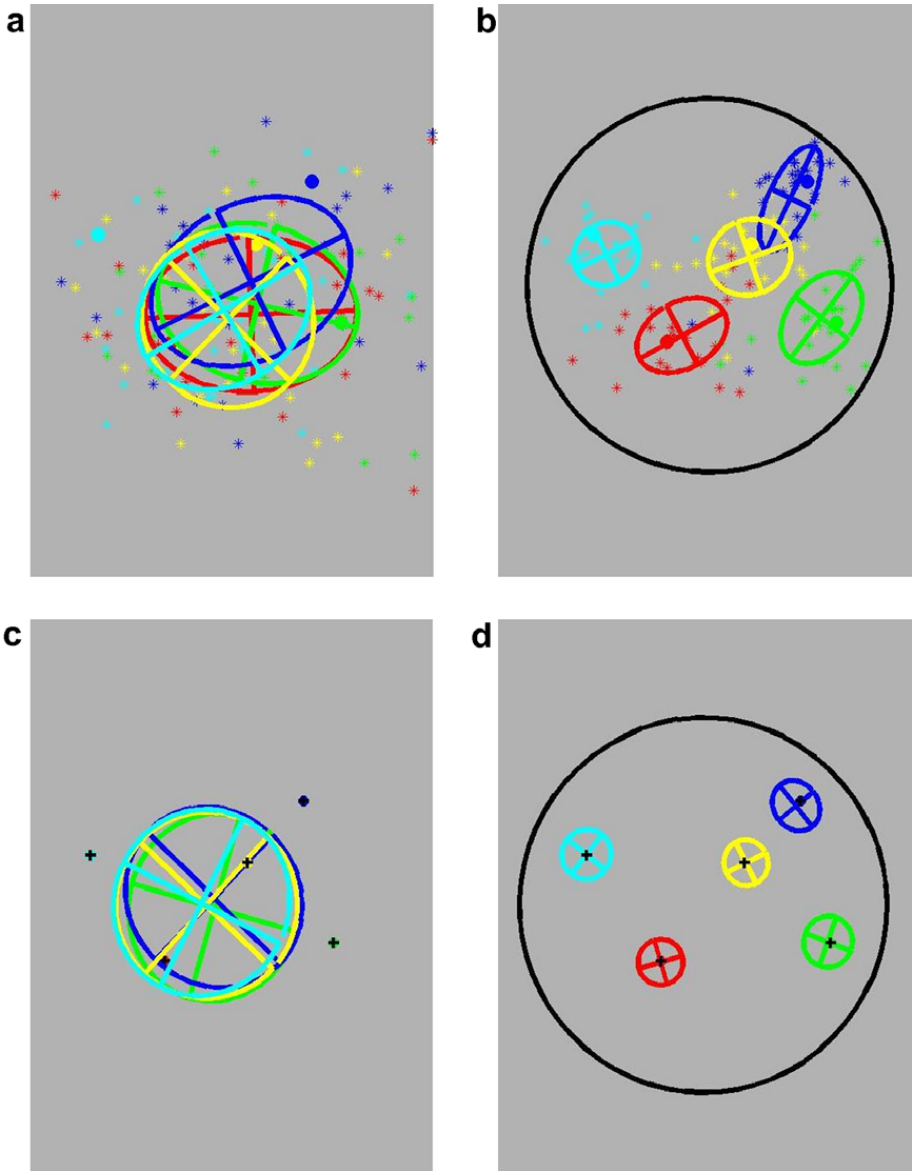


Fig. 14. Endpoints of the trajectories in Experiment 2, circular room. The black or heavy coloured symbols mark the goal locations. (a and b) Experimental data (scatter-plot, means and error ellipses) (c and d): simulation (only means and error ellipses are shown), Left column (c and d: the XL environment (circular wall outside tracked area). Right column (b and d): S-environment. As predicted by the model, the variance ellipses of the empirical data are smaller in smaller rooms.

Author's personal copy

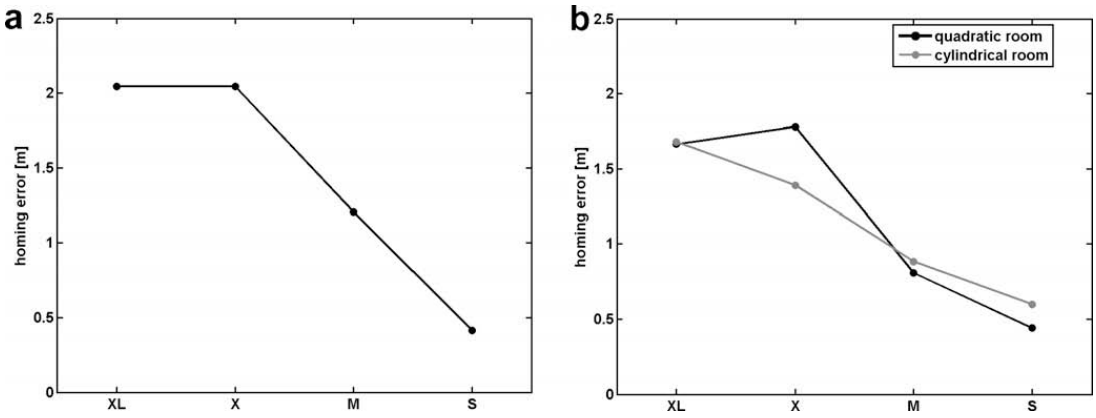


Fig. 15. Homing error in rooms of different size and shape. (a) Simulation for the circular room. (b) Empirical results. Reduction of the environment size results in an improved homing accuracy.

conditions. The quantitative prediction for the radial and tangential diameters of the confusion areas (Fig. 12b) is not matched, but the parallelity of the two diameter curves is clearly visible in Fig. 12a. This parallelity is not predicted by any of the other models (cf. Table 1).

4.2. Models based on wall-distances

If wall-distance was used as a cue to place localization, the radial diameter of the confusion area will depend only of the quality of this distance measurement and will therefore not be affected by colour modulation. Still, in the closest-wall-segment and boundary-vector models, the size of the confusion area will increase for low colour modulation and larger rooms, but only in the tangential direction. As a result, the shape of the confusion area at least for peripheral locations should vary, from circular to elongated or compressed with respect to the radial axis. This prediction cannot be confirmed. Indeed, the shape of the confusion areas was roughly circular in all conditions. The wall-distance models are thus consistent with the results presented in Fig. 11 (homing error decreases with colour modulation) and Fig. 15 (homing error increases with room size), but not with the overall circular shape of the confusion areas (Fig. 10a, b and Fig. 12).

The decrease of homing accuracy with colour modulation is apparent for the peripheral goals, but not for the central ones (Fig. 11b). For the central goals, information from image differences is less reliable, due to a reduced local image variance (Hübner & Mallot, 2007; Stürzl & Mallot, 2006). On the other hand, a simple wall distance strategy works better for central goals, since the area of the equidistant ring gets smaller for more central point. One might therefore speculate that subjects use different strategies for different goal eccentricities, image difference close to the wall and "thigmotaxis" (Jacobs et al., 1998; Kallai et al., 2007) in the centre.

How can subjects judge the distance to the walls? The strongest cue is probably the elevation of the wall in the visual field together with the vertical visual angle subtended by it. Another possibility is motion parallax obtained from the high-contrast upper and lower margins of the wall. Stereocues were absent in our setup since both eyes were presented with the same image. Motion parallax may provide additional information if movements in the radial direction are carried out. Note that the elevation cue to distance also enters our intensity-based image comparison scheme, since comparisons are done for full two-dimensional images, not just for horizontal rings (see also Stürzl et al., 2008 for edge information in image difference models). Still, the algorithm does not predict the different contrast-dependencies for central and peripheral points. This further supports the idea that a separate, distance based strategy is adopted at the central points.

4.3. Initial approach direction

The initial approach directions chosen by the subjects when first starting their walk to the goal show fairly small deviations from the ideal direction. This is apparent from the sample trajectories shown in Fig. 7 but also from the quantitative plots of Fig. 9. When judging the numbers, one should keep in mind that all deviations below 90° will lead to a reduction of the distance to the goal. For the head directions shown in Fig. 8, a similar trend is obvious at least for the small room (Fig. 8a): here the subject is looking predominantly towards the goal location. Taken together, this findings suggest that subjects are not only able to recognize the goal location once they reach it, but that they also are able to infer the appropriate approach direction from a distance. From the discussion of homing accuracy given above, it seems clear that this performance must be based on the same image cues as place recognition itself. Indeed, the visual homing algorithms from the insect and robotics literature (Cartwright & Collett, 1983; Franz et al., 1998; Vardy & Möller, 2005) solve both problems simultaneously. In the "optic flow"-approach of Cartwright and Collett (1983) and Vardy and Möller (2005), feature points from the stored and current images are matched and used to calculate an optic flow field whose focus of expansion will point to the target location. In the "image warping" approach of Franz et al. (1998), predictions of expected image change are produced for a number of directions and the predicted images are compared to the target image. The travel is continued in the direction associated with the biggest improvement in image similarity.

Author's personal copy

From the data presented in this paper, we cannot decide which of the two algorithms might be used in human navigation.

4.4. Landmarks and objects

Our results add further evidence to the distinction between objects and landmarks discussed already in the introduction section. In a domain-specific view of cognition (for review, see Kinzler & Spelke, 2007), spatial cognition is usually treaded as one subpart of cognition which to some extend is independent of visual and object cognition. Biologically, this makes sense because navigational abilities are useful for all freely moving animals while object cognition seems most useful only for those who already evolved hands (or beaks) for manipulation. The type of landmark navigation demonstrated in this paper is not dependent on a powerful system for object recognition. It might therefore be part of an evolutionarily old system operating already before the invention of object recognition proper.

5. Conclusion

In summary, our results indicate that humans are able to approach to and recognize places based on raw image information void of localized cues, landmarks, or objects. This performance is similar to visual homing behaviour described for various insects. As has been shown in a large number of robotic studies, image-based homing is a powerful strategy in real-world navigation. Our findings do not rule out other forms of landmark usage such as localization with respect to local wall-distance or recognized objects, which may be used additionally in richer environments.

Acknowledgements

Supported by the Deutsche Forschungsgemeinschaft (Grants Gi373/1-1 and SFB 550) and by the European Commission (6th FP NEST Pathfinder Project "Wayfinding"). We are grateful to the Max-Plank-Institute for Biological Cybernetics, Tübingen (Prof. Bülthoff), for providing the HMD used in the experiments and to Jan Burk for collecting part of the data.

References

Aginsky, V., Harris, C., Rensik, R., & Beusmans, J. (1997). Two strategies of learning a route in a driving simulator. *Journal of Environmental Psychology, 17*, 317–331.

Allen, G., Kirasic, K., Siegel, A., & Herman, J. (1979). Developmental issues in cognitive mapping: The selection and utilization of environmental landmarks. *Child Development, 50*, 1062–1070.

Appleyard, D. (1969). Why buildings are known. *Environment and Behavior, 1*, 131–156.

Barry, C., & Burgess, N. (2007). Learning in a geometric model of place cell firing. *Hippocampus, 17*, 786–800.

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., et al (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences, 17*, 71–79.

Campbell, F., & Maffei, L. (1974). Contrast and spatial frequency. *Scientific American, 231*, 106–114.

Cartwright, B., & Collett, T. (1983). Landmark learning in bees. Experiments and models. *Journal of Comparative Physiology A, 151*, 521–543.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition, 23*, 149–178.

Cornell, E., Heth, C., & Broda, L. (1989). Children's way finding: Response to instructions to use environmental landmarks. *Developmental Psychology, 25*(5), 755–764.

Daniel, M. P., Tom, A., Manghi, E., & Denis, M. (2003). Testing the value of route directions through navigational performance. *Spatial Cognition and Computation, 3*(4), 269–289.

Durier, V., Graham, P., & Collett, T. S. (2003). Snapshot memories and landmark guidance in wood ants. *Current Biology, 13*, 1614–1618.

D'Zmura, M., & Singer, B. (1999). Contrast gain control. In K. Gegenfurtner & L. Sharpe (Eds.), *Color vision. From genes to perception* (pp. 369–385). Cambridge UK: Cambridge University Pres.

Franz, M., Schölkopf, B., Mallot, H. A., & Bülthoff, H. H. (1998). Where did I take that snapshot? Scene-based homing by image matching. *Biological Cybernetics, 79*, 191–202.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Graham, M., Good, M. A., McGregor, A., & Pearce, M. (2006). Spatial learning based on the shape of the environment is influenced by properties of the objects forming the shape. *Journal of Experimental Psychology: Animal Behavior Processes, 32*, 44–59.

Hauser, K. (2005). Erzeugung homogener Farbverläufe anhand psychophysischer Messungen am Menschen (Generation of homogenous color gradients based on psychophysical experiments with humans). Master's thesis, University of Tübingen, Germany.

Heft, H. (1979). The role of environmental features in route-learning: Two exploratory studies of way finding. *Environmental Psychology and Nonverbal Behaviour, 3*, 172–185.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature, 370*, 57–59.

Hermer-Vasquez, L., & Spelke, E. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology, 39*, 3–36.

Hübner, W., & Mallot, H. A. (2007). Metric embedding of view graphs. A vision and odometry-based approach to cognitive mapping. *Autonomous Robots, 23*, 183–196.

Janzen, G., Hermann, T., Katz, S., & Schweizer, K. (2000). Oblique angled intersections and barriers: Navigating through a virtual maze. *Lecture Notes in Artificial Intelligence, 1849*, 277–294.

Janzen, G. (2006). Memory for object location and route direction in virtual large-scale space. *The Quarterly Journal of Experimental Psychology, 59*(3), 493–508.

Janzen, G., & van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience, 7*(4), 673–677.

Jacobs, W. J., Thomas, K. G., Laurance, H. E., & Nadel, L. (1998). Place learning in virtual space II. Topographical relations as one dimension of stimulus control. *Learning and Motivation, 29*, 288–308.

Kallai, J., Makany, T., Csatho, A., Karadi, K., Horvath, D., Kovacs-Labadi, B., et al (2007). Cognitive and affective aspects of thigmotaxis strategy in humans. *Behavioural Neuroscience, 121*, 21–30.

Kinzler, K. D., & Spelke, E. S. (2007). Core systems in human cognition. *From Action to Cognition, 164*, 257–264.

Learmonth, A., & Nadel, L. (2002). Toddler's use of metric information and landmarks to reorient. *Journal of Experimental Child Psychology, 80*, 225–244.

Lynch, K. (1960). *The image of the city*. Cambridge, MA: MIT Press.

Mallot, H. A., & Gillner, S. (2000). Route navigation without place recognition: What is recognized in recognition-triggered responses? *Perception, 29*, 43–55.

Milliken, G. A., & Johnson, D. E. (1992). Analysis of messy data. In *Designed experiments* (Vol. I). Boca Raton, Fl: Chapman & Hall/CRC.

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

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the cognitive map. *Nature Reviews Neuroscience, 7*, 663–678.

Restat, J. D., Steck, S. D., Mochnatzki, H. F., & Mallot, H. A. (2004). Geographical slant facilitates navigation and orientation in virtual environments. *Perception, 33*(6), 667–687.

Siegel, A. W., & White, S. (1975). The development of spatial representations of large scale environments. In H. W. Reese (Ed.). *Advances in child development and behavior* (Vol. 10, pp. 10–55). New York: Academic Press.

Stankiewicz, B. J., & Kalia, A. (2007). Acquisition and retention of structural versus object landmark knowledge when navigating

Author's personal copy

through a large-scale space. *Journal of Experimental Psychology – Human Perception and Performance, 33*, 378–390.

Stürzl, W., Cheung, A., Chen, K., & Zeil, J. (2008). The information content of panoramic images I: The rotational errors and the similarity of views in rectangular experimental arenas. *Journal of Experimental Psychology: Animal Behavior Processes, 34*, 1–14.

Stürzl, W., & Mallot, H. A. (2002). Vision-based homing with a panoramic stereo sensor. *Lecture Notes in Computer Science, 2525*, 620–628.

Stürzl, W., & Mallot, H. A. (2006). Efficient visual homing based on Fourier transformed panoramic images. *Robotics and Autonomous Systems, 54*, 300–313.

Stürzl, W., & Zeil, J. (2007). Depth, contrast and view-based homing in outdoor scenes. *Biological Cybernetics, 96*, 519–531.

Tinbergen, N., & Kruyt, W. (1938). Über die Orientierung des Bienenwolfes (*Philanthus triangulum* Fabr.) III: Die Bevorzugung bestimmter Wegmarken. *Zeitschrift für vergleichende Physiologie, 25*, 229–344.

Trullier, O., Wiener, S., Berthoz, A., & Meyer, J.-A. (1997). Biologically based artificial navigation systems: Review and prospects. *Progress in Neurobiology, 51*, 483–544.

Vardy, A., & Möller, R. (2005). Biologically plausible visual homing methods based on optical flow techniques. *Connection Science, 17*, 47–89.

Waller, D., & Lippa, Y. (2007). Landmarks as beacons and associative cues: Their role in route learning. *Memory and Cognition, 35*, 910–924.

Wiener, J. M., Franz, G., Rossmanith, N., Reichelt, A., Mallot, H. A., & Bülthoff, H. H. (2007). Insovist analysis captures properties of space relevant for locomotion and experience. *Perception, 36*, 1066–1083.

Zeil, J., Hofmann, M., & Chahl, J. (2003). Catchment areas of panoramic snapshots in outdoor scenes. *Journal of the Optical Society of America A, 20*, 450–469.

Zugaro, M. B., Arleo, A., Dejean, C., Burguiere, E., Khamassi, M., & Wiener, S. I. (2004). Rat anterodorsal thalamic head direction neurons depend upon dynamic visual signals to select anchoring landmark cues. *European Journal of Neuroscience, 20*, 530–536.