Embodied Spatial Cognition: Biological and Artificial Systems*

Hanspeter A. Mallot and Kai Basten
Department of Cognitive Neurobiology
Faculty of Biology, University of Tübingen, Germany
hanspeter.mallot@uni-tuebingen.de

Abstract

In this paper we sketch out a computational theory of spatial cognition motivated by navigational behaviours, ecological requirements, and neural mechanisms as identified in animals and man. Spatial cognition is considered in the context of a cognitive agent built around the action-perception cycle. Besides sensors and effectors, the agent comprises multiple memory structures including a working memory and a longterm memory stage. Spatial longterm memory is modeled along the graph approach, treating recognizable places or poses as nodes and navigational actions as links. Models of working memory and its interaction with reference memory are discussed. The model provides an overall framework of spatial cognition which can be adapted to model different levels of behavioural complexity as well as interactions between working and longterm memory. A number of design questions for building cognitive robots are derived from comparison with biological systems and discussed in the paper.

1 Introduction

1.1 Embodied Spatial Cognition

Behaviour in animals and man can be described by the perception-action-cycle linking sensory and motor components via central processing within the agent and via a feedback loop established by the environment. The cognitive processing required to initiate and control higher level behaviour will thus depend on the available sensorium, the effectors, the environment, and the behavioural tasks pursued by the agent. A major difference between robotic and biological systems lies in their respective sensors and the types of data these sensors provide. In biological systems, local views

*In print in *Image and Vision Computing* 27:1658-1670, 2009

of the environment and distance measurements obtained from egomotion are prominent types of sensory data, whereas range and absolute position sensors (GPS) play a large role in technical systems. In biological systems, the association of a recognized view with some motor action (recognition triggered response), i.e. a stereotyped perception-action-pair, is therefore considered a basic element of spatial memory which at the same time allows to generate simple route following behaviour. More abstract forms of spatial memory are required for the planning of novel routes (cognitive maps) indicating that the abstraction level of memory scales with the complexity of the tasks performed.

In this paper, we will develop a framework for spatial cognition based on the action-perception-cycle and a biologically plausible sensorium. The resulting theory of spatial cognition may be called *embodied* in the sense that cognitive processing is shaped by the sensorium, the effectors and the behavioural tasks to be carried out by the agent. Following the discussions presented by Lakoff (1987) [49] and Gallagher (2005) [25], the following aspects of embodiment will be considered:

1. Adaptation to sensorium: Perception and the available types of sensory information influence the computations to be performed by the central processor.

2. Abstraction: Memory structures are "close" to the data structures of sensors and motor control, i.e. the transformations from sensors to memory and memory to action require only a small number of processing steps.

3. Task-specificity: Central processing is adapted to the behavioural repertoire determined by motor abilities and tasks.

4. Incrementality: Learning of spatial representations occurs incremental both in a local sense (learn about current place) and in a tasks-specificity sense (learn how to solve current task).

As a consequence of the task-specificity, the cognitive apparatus of animals and man is not a universal computer, but rather a collection of partially independent adaptations to various problems which are relevant in an animal's life (see, for example, Hauser 2000 [33]). Indeed, the now common delineation of *spatial* cognition from other domains such as visual cognition, tool use, and social cognition (e.g., Spelke & Kinzler, 2007) [73] is an example for the task-specific approach. In this paper, we limit ourselves to the spatial domain, which is also the most widespread domain of cognition in the animal kingdom. Possible extensions of the framework to other domains are left to future research.

The framework for spatial cognition presented in this paper starts from stimulus-response schemata and simple versions of the perception-action-cycle. The cognitive level is reached by including various types of graph structures in the central memory structures allowing to model

topological navigation and declarative memory[1]. The framework is extended to include hierarchical representations, route planning, functionality of places, and metric information. The approach is evolutionary in that the adaptive value of a behavioural competence and the required information processing are scaled up jointly, in small, simultaneous steps.

The main scope of this paper is to describe a general framework for modeling biological navigation systems in various species. The framework also allows comparisons with technical approaches and may be useful in situations such as navigation of large outdoor environments, learning high level spatial maps representing regions and places functionalities, and interfacing navigation systems with human users.

Table 1: Task hierarchy in spatial cognition. S: Stimulus, R: Response

Task	Required information processing and memory
recognizing places	memory of local position information characteristic of places
finding home after excursion	two mechanisms: landmark guidance and/or path integration
following a route	associate places with motor actions (S-R or S-R-S)
recombining known route segments	graphs (networks) of S-R-S associations
route selection and planning	working memory
cross-country shortcuts	metric embedding of places
communicating about space	naming places and actions

1.2 Tasks in spatial cognition: Repertoire and ontologies

As a result of its evolutionary origin, the ontology of spatial knowledge, i.e. the selection of data types represented in spatial memory, reflects the behavioural repertoire for which it evolved. Table 1 gives an overview of tasks or behavioural competences in spatial cognition together with the required representations and information processing abilities, which may be summarized as the spatial ontology. While the main ordering criterion is based on behaviour, not on information

[1] Declarative memory is a part of longterm memory storing declarations rather than procedures or associations. In spatial memory, map knowledge is considered declarative while route knowledge is not.

processing, the resulting hierarchy is similar to the hierarchies proposed by Kuipers (1978, 2000) [46, 47] and by Trullier et al. (1997) [80].

Place recognition and homing. The simplest tasks requiring some sort of spatial memory occur in a life style called *central place foraging*, found in many insects or other animals building some sort of borrow or nest and taking foraging trips from there (e.g., Wehner 2002) [84]. One type of memory needed is a longterm or reference memory of the place itself, i.e. a representation of sensory cues allowing to recognize the place once it is reached. Depending on this recognition mechanism, there will always be a neighbourhood around that place from which homing is possible (catchment area) and a smaller neighbourhood within which no further approach can be made (confusion area), see Franz et al. (1998) [23]. In more complex memories, the mechanisms developed for the actual home may be applied to additional places in just the same way.

Another strategy that may be used for homing especially if excursions are longer, is path integration. It is based on continuous updating of a so-called home-vector, i.e. a metric representation of the vector linking home to the agent's current position. This updating requires some kind of working memory since the home-vector changes with egomotion. Vector knowledge may also be stored in longterm memory, for instance in the form of labels attached to remembered places and specifying the distance and direction to the central place (e.g., Cartwright & Collett 1987) [11].

Route following. Extending the "catchment area" of the central place by storing additional places as stepping stones leads on to the next behavioural level which is following a route or sequence of steps. Routes are generally thought to be composed of discrete decision points which have to be recognized from some longterm memory contents, and route segments which may be traveled without longterm memory involvement. Decisions at the decision points are based on stimulus-response (SR) learning achieved by operant conditioning or reinforcement learning. While the chain-idea of route memory is theoretically appealing, direct evidence for these view comes mostly from maze experiments, where decision points are defined by the experimenter (e.g., Mallot & Gillner 2000, Janzen & van Tourennout 2004, Waller & Lippa 2007) [51, 42, 83]. Of course, routes are also formed in cluttered, irregular environments where no explicit decision points are provided (Hurlebaus et al., 2008) [40]. Route knowledge in open environments has been convincingly demonstrated e.g. in ants (Wehner et al. 2006) [85] and bats (Schaub & Schnitzler, 2007) [67]. In these cases, routes seem to be continuous, giving no evidence for the discrete, stepwise character assumed in standard models of route memory. Still, the discrete structure may underlie also continuous routes, if decision points are close and catchment areas are large.

Inferring novel routes. The map level of spatial knowledge is reached if navigators are able to recombine parts of known routes into novel routes (Tolman 1948, O'Keefe & Nadel 1978) [79, 62].

Extending the route concept to maps amounts to replacing the chains of decision points and actions by networks or graphs which contain bifurcations and loops. In this situation, actions cannot simply be associated to decision points, since one decision point may indeed allow different actions such as "turn left" or "move uphill". Choosing a particular action depends on the currently pursued goal and requires two novel abilities: (i) a so-called declarative knowledge of the type "choosing action A_i at place P_j will lead to place P_k" (a three-way association: stimulus, response, expected next stimulus [S-R-S']) and (ii) a planning stage allowing to select a sequence of places eventually leading to the goal. The planning stage is a type of spatial working memory generating executable routes from the map-like long-term memory. It may also allow to consider alternative routes and make high-level navigational decisions. The distinction between routes and maps, originally suggested with theoretical and behavioural arguments, is now well established also on neurophysiological grounds. Hartley et al. (2003) [32], for example, have shown that the basal ganglia are involved in stereotyped route following while map navigation activated cortical areas.

The process of generating a navigable route from a cognitive map[2], i.e. route planning, requires a working memory to store the steps of the route while navigating. Bumble bees, for example, show systematic "trap-line" foraging patterns in spatial arrays of feeders, requiring both route memory and some route planning abilities (Ohashi et al. 2007) [60]. In humans, path planning is generally studied in the context of choosing between alternative routes to the same goal (Bailenson et al. 2000, Wiener & Mallot 2003, Wiener et al. 2004) [6, 88, 90]. We will discuss these performances in more detail below.

Metric knowledge. Metric information may be used for quite different tasks in spatial cognition. Motor activity always includes some metric knowledge about the movements performed. Together with egomotion estimates from visual and vestibular senses, this information is used in path integration. In longterm-memory, metric information is important to find shortcuts over novel terrain, i.e. shortcuts which are not just combinations of known route segments. Also, it will help making choices between alternative routes in route planning. Theoretically, three possible types of metric long-term memory may be distinguished, (i) full metric maps, (ii) metric embeddings, and (iii) local metric information.

In the *full metric map* approach, the cognitive map is a continuous chart in which known places or objects are plotted. Each location is represented in the full metric map (with some resolution) and a label provides information whether this location is occupied or not. Alternatively, the graph of places and actions discussed in the last paragraph might be augmented by *local metric information* such as distances associated with graph links or angles associated to adjacent

[2]The term cognitive map is used for different types of declarative spatial memory. In the simplest case, it is a graph of places and place transitions, but it may also include metric and hierarchical structure. We will use the term only in contexts where the different meanings cannot be confused.

pairs of graph links. Local metric information is relational and does not require the specification of a coordinate system. An intermediate case is *metric embedding* where coordinate values are associated to nodes only. Like the full metric map, metric embedding requires a coordinate system but does not represent points which are not actually occupied by a place node of the graph.

While these theoretical concepts are quite clear, experimental evidence on metric long-term memory is scarce. Metric judgments often lack symmetry, i.e., distance from A to B may be judged differently from the distance from B to A, (McNamara & Diwadkar, 1997) [52] or violate the triangular inequality. Foo et al. (2005) [20] have adapted the standard triangular completion paradigm used in studies of path integration (a working memory task) for long term memory. If two legs of a triangle are learned independently until each leg is well represented in longterm memory, the pointing performance between the open ends of the two legs is still poor. This is surprising if the places (vertices of the triangle) were represented with coordinate values in some global coordinate frame. Similarly, if subjects are taken to an unknown environment and are given repeated training for 10 subsequent days, their performance in pointing tasks between various points on their daily route is poor and does not improve with training (Ishikawa & Montello 2006) [41]. Again, this result is surprising if metric localization in a global framework is assumed (Gallistel 1990) [26].

Communication about space. In human spatial cognition, navigational aids such as drawn or printed maps, verbal directions, street signs and public transportation systems play an essential role. These navigational aids depend on social interaction and language. The behaviours which they subserve are therefore not purely spatial, but integrate elements from the spatial and social domains.

As a result of the predominance of such navigational aids in human spatial cognition, it may appear that spatial behaviour in general relies on language-based mechanisms. The view taken in this paper is rather opposite. Since wayfinding abilities can be observed throughout the animal kingdom, they must be a forerunner of language based cognition, not a consequence. Indeed, in their review on language evolution, Hauser et al. (2002) [34] see spatial cognition as part of a general "concept system" which is a preadaptation to language. This does not preclude the possibility of closer interactions between language and space in humans.

In accordance with this view, navigational behaviour and social communication in animals are rather independent. While navigational competences are widespread, communication abilities are rather rare and often limited to mating and warning. The most sophisticated example for communicating spatial information is certainly the honeybee dance language discovered originally by von Frisch [82]. Here, a honeybee returning to its hive from a profitable food source communicates the distance and direction of the food location to its nestmates. The key feature of this communica-

tion, however, is recruitment, i.e. getting fellow worker bees to search in the advertised location. Distance and direction to the food source are encoded in the frequency of "waggle"-movements of the abdomen and dance direction relative to the gravitational vertical, respectively. The function of these spatial codes is purely for communication, they are not the information that the dancer would use to go back to the food source.

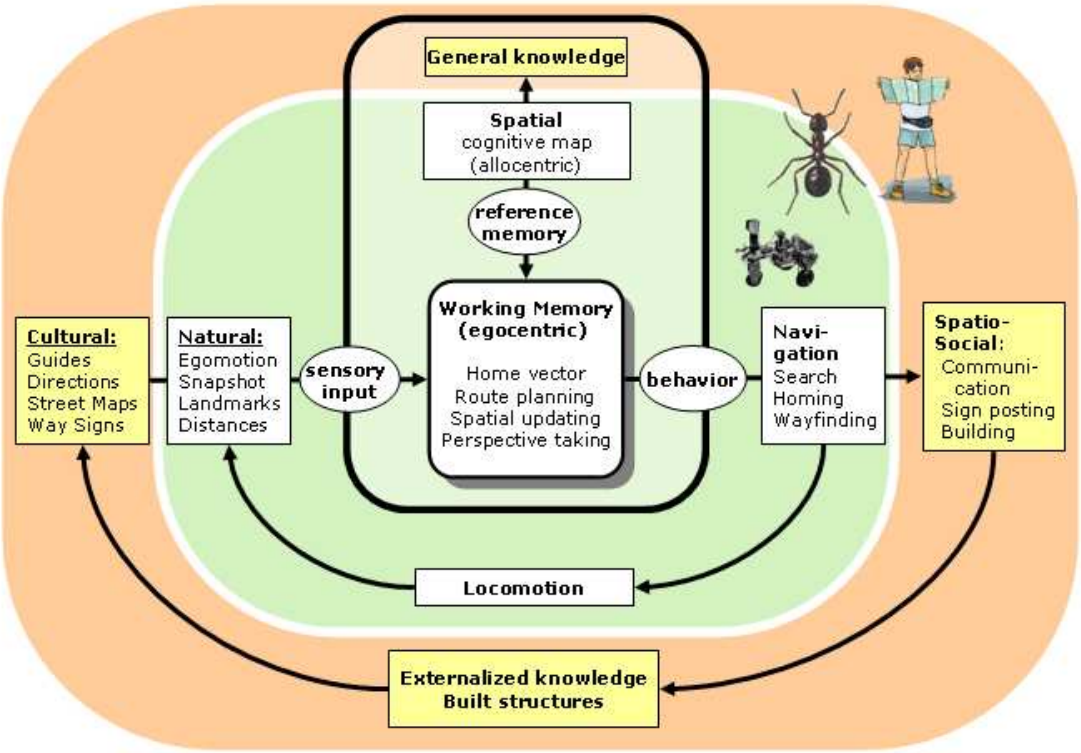

Figure 1: Cognitive agent for spatial behaviour. Inner box with heavy outline: Agent with input and output. Inner shell: wayfinding behaviour, as is found in animals and humans. Outer shell: Interactions of social and spatial behaviour restricted mostly to humans. To date, robots generally also operate in the wayfinding domain.

1.3 A Cognitive Agent

Animal behaviour is generally analysed in the perception-action-cycle, where central processing is considered as some kind of controler trying to keep the organism in a favourable state. For the discussion of spatial cognition, it is important to distinguish within this controler different types of memory, including working and longterm (or reference) memory components. In order to structure the discussion, we briefly discuss a cognitive agent for spatial tasks depicted in Fig. 1. Clearly, the involvement of the various memory components will vary for different navigational

mechanisms and different animal species. The similarity to the standard von Neuman architecture of a computer is not intended to mean that parallel processing is considered unlikely. Indeed, the working memory stage is a rich, structured system that contains a planning device selecting between different subgoals and navigational strategies. This planing device might well realize a subsumption architecture. Also, it should be stated that the boxes making up this agent are only logical in nature, giving no one-to-one relation to parts of the brain.

In order to address different types of spatial behaviour in robots, animals, and humans, we have divided the environment in two parts. The inner shell in Fig. 1 includes natural environments not intentionally changed by activities of the agent or conspecifics. Here, the standard wayfinding behaviour takes place as discussed e.g. by Trullier et al. (1997) [80]. Humans (and some animals) change their environments by building structures or by providing information to conspecifics. The according behaviours require additions in various parts of the agent as will be discussed below.

Sensory input. In the wayfinding shell, sensory input relevant to spatial behaviour includes two types, landmarks (or local position information) and egomotion. For both types, the visual modality will be of great relevance.

Sensory input relevant to spatial tasks can be described by the local position information, i.e. the total sensor readings obtained from each position. Local position information can be thought of as landmark information in a very broad sense, which may be transformed to more explicit landmark information by applying different amounts of processing, leading (in the visual domain) to snapshots, geometric descriptions, or recognition of landmark objects. A second class of sensory information concerns egomotion which may be measured through different sensory modalities. Together, landmark and egomotion cues comprise "natural" cues whose presence in the environment is incidental and not due to some sort of information posting. In contrast, information is said to be "cultural", if available due to according efforts of the navigator or its conspecifics. In animals, examples for provided information include spatial communication such as the honeybee waggle dance, pheromone trails in ants, or olfactory territory markings. In humans, such provided information is abundant, including verbal directions, streets and corridors, way signs, printed maps, etc.

Reference memory. Longterm memory of space has been the subject of research on spatial cognition for decades. The emerging view of the "ontology" of spatial memory, i.e. of the types of structures represented, includes places (represented e.g. as feature vectors of local position information), actions required to move from one such place to another, local metric information associated to such actions, etc. In many cases, spatial performance will additionally rely on general knowledge not referring to a particular place. Examples range from simple rules like "walk downhill

to reach the water" to general knowledge about buildings or the abilities needed to interpret a printed map.

Working memory. Baddeley (1986, p. 34) [5] defined *general working memory* as the "temporary storage of information that is being processed in any of a range of cognitive tasks". In this sense, it is the central processing unit integrating sensory data and memory retrieved from the reference memory stage to control behaviour and generate new reference memory items. Working memory tasks in animal spatial cognition include simple path integration based on a home-vector representation, route planning, and object permanence (spatial updating). The spatial framework used in these tasks is ego-centric. Although working memory is depicted as one central stage in Fig. 1, we do not exclude the possibility, that different tasks have separate working memory instances.

Behaviour and action. The "core" spatial behaviour studied extensively in both animals and humans is wayfinding, i.e. getting from one place to another. Indeed, the wayfinding hierarchy sketched out in Section 1.2 is usually considered to cover the entire field of spatial cognition, at least in animals. However, wayfinding is not the only way in which both humans and animals interact with space. Building (nest) structures for individuals or groups, establishing and defending territories, or cooperation in larger groups and swarms all require some spatial abilities.

Environment. In the general view of spatial behaviour outlined here, the environment has to be considered as a part of the entire system. Human navigation rarely takes place in the wild, but deals mostly with finding intersections, reading signs and maps, or following verbal instructions. Descriptions of man-made environments together with their natural origins are sought in spatial information theory and GIS.

1.4 Design questions for spatial cognitive agents

For the spatial domain of cognition, a computational theory is emerging which builds on a number of simple mechanisms of orientation and place recognition which are integrated and scaled to an overall system. By and large, the resulting theory extends Kuipers' (1978) [46] spatial semantic hierarchy, which bases longterm, declarative knowledge of space (the cognitive map) on a graph of recognizable places and control rules describing transition between these places. In Section 2 we will review the current state of the graph theory of cognitive maps theory from a biological point of view.

From the comparison of biological and robotic spatial cognition systems, a number of design questions can be identified which are treated differently in different approaches. These design

questions include

1. Landmark representation: Are landmarks a separate concept or data type included in the cognitive map or are they just properties of places from which they are perceived?

2. Metric information: While metric is clearly included in path integration and other motor mechanisms it is not clear how much metric information is actually represented in the cognitive map.

3. Hierarchies: The concept of a "place", even if envisaged as a geometric locus, will always include some spatial extension, if only as a result of uncertainty. In general, however, spatial knowledge will also represent regions of various hierarchical levels, which may also overlap.

4. Functionality: Spatial knowledge may be represented jointly with non-spatial information. As an example, consider the world graph approach of Arbib & Lieblich (1977) [3] where possible behaviours are stored together with each place representation.

5. Using the cognitive map: Reading the cognitive map requires a planning stage which buffers the information relevant for the current task and translates it into subgoals and motor actions. We will discuss the representation of actions as labels attached to graph links, and the planning stage as a spatial working memory.

The paper is structured by the graph approach to spatial cognition, which has been briefly discussed already in Section 1.2. In Section 2, we will consider various types and topologies of graphs that have been used for modeling spatial cognition. In Sections 3 and 4, individual components of the graph will be discussed. We will finish with some considerations concerning multiple memories of space, such as reference and working memories.

2 Graph representations of spatial memory and spatial behaviour

2.1 General

As usual, we consider graphs as structures composed of a set of nodes (or vertices) $V = \{v_i, i = 1, ..., n\}$ and a set of directional links (or edges) $E = \{e_i = (v_{i1}, v_{i2}, \lambda_i), i = 1, ..., m\}$. In our context, nodes are interpreted as spatial entities such as places, poses (place plus viewing direction), regions, etc. They are treated as empty containers that may carry different types of information such as views visible from each node, coordinates, etc. We will call these types of information "labels". The links are treated as triples $e_i = (v_{i1}, v_{i2}, \lambda_i)$ where $v_{i1}, v_{i2} \in V$ are the nodes

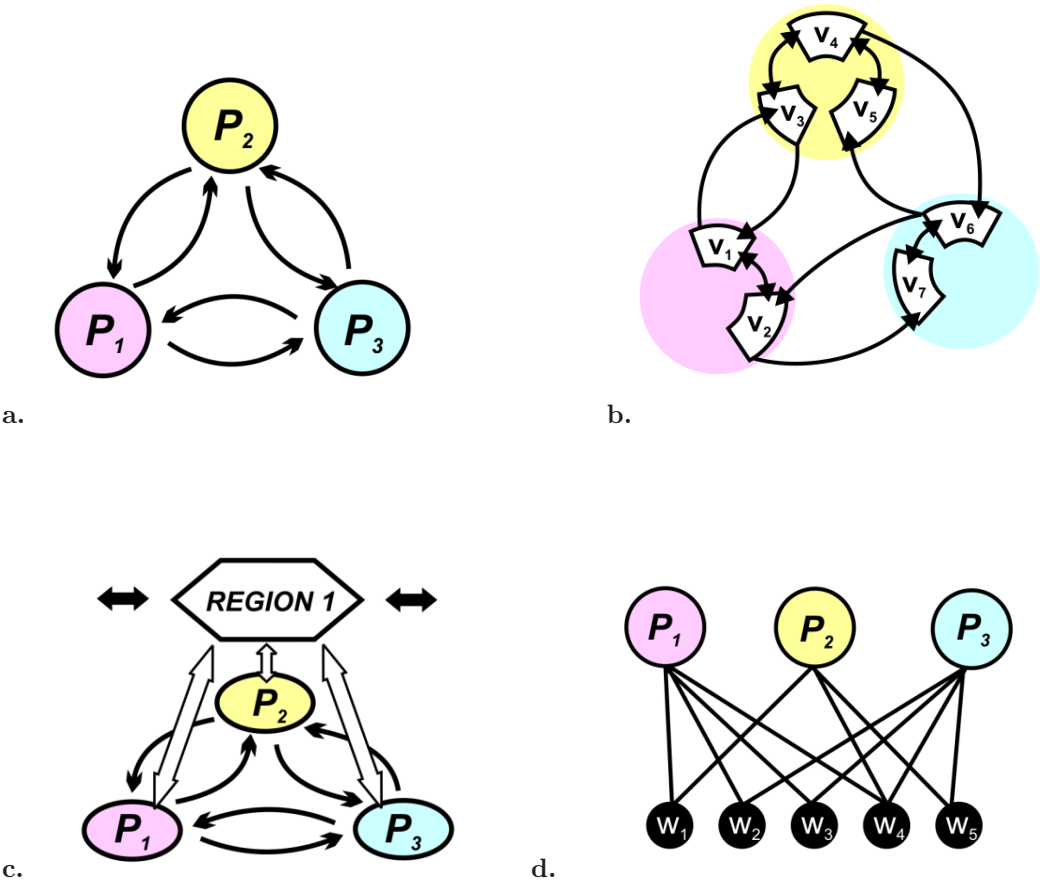

Figure 2: Graph models of spatial memory. **a.** Place graph with nodes (places) P_i. **b.** View graph with nodes v_i. The coloured regions indicate viewpoints (places) from which the views were taken. **c.** Hierarchical graph with place nodes P_i and an additional regional node. **d.** Bipartite Graph of places P_i and ways W_i.

connected by the link and λ_i is a label or set of labels attached to the link. These labels may contain information needed to proceed from the start node to the end-node of the link.

Graph descriptions of space are used extensively in computational geometry, in particular in the context of architectural planning. The scope of the discussion here is not so much on space itself, but on the mental representation of space. Still, many concepts will be reminiscent of ideas used in computational geometry.

Historically, graphs have been used to model spatial memory and spatial behaviour at least since the work of Tolman (1932, Fig. 61, p. 177) [78]. In Tolman's terms, the nodes of the graph are goals, or "ends" and the edges are "means-ends"-associations; the whole network is described as a "means-ends-field". Clearly, this is a very general characterization, intended to describe behaviour and problem solving in general, rather than a pure representation of space. An end,

in Tolman's scheme, is a state taken by the agent and may be both spatial (e.g., "I have reached place A") or non-spatial (e.g., "I have obtained object B"). A means is a procedure for achieving state transitions. Thus, the means-ends-field appears quite similar to the idea of a finite state machine and its associated graph of state transitions.

More recent graph models of spatial memory and spatial behaviour have used different specifications of what nodes and links of the graph actually are or represent. Choices include places, views, ways, and regions for the nodes and according transition rules for the links. In this section, we discuss the main types of "space graphs", i.e. graph structures used in modeling spatial cognition; an overview is given in Fig. 2. We do not consider here the recognition of the various nodes or the information used for this task; this question will be postponed to Section 4.1.

2.2 Topologies

One dimension, along which space graphs may vary is topology. In animal behaviour, three major types of topologies are generally considered: chains, stars, and nets.

Chains of places and place transitions are called route memories. It is generally assumed that the recognition of a place (or other type of graph node; see below) triggers some action which then leads to the next node in the chain ("recognition-triggered response" of Trullier et al., 1997 [80]; "direction" of O'Keefe and Nadel, 1978 [62]). Thus, the graph nodes have a label saying what to do when reaching the node (see below, Section 4).

Star topologies are often considered for insect spatial memories, especially for honeybees and ants. The centre of the star is the home or hive, or more generally the central place of a central place forager (Cartwright & Collett 1987, Menzel et al. 1996, Dyer 1996, Wehner et al. 2006) [11, 55, 18, 85]. Honeybees are able to learn routes from the hive to various feeders and back, but do not seem to take shortcuts between different feeders with non-overlapping catchment areas. In cases where shortcuts have been observed, conspicuous landmarks may have been used that result in large catchment areas. It should be noted, however, that experiments using radar techniques for tracking whole flight paths of bees, seem to indicate that honeybees might be able to integrate at least two vectors and to choose between at least two locations (Menzel et al. 2005, 2006) [56, 57].

True network topologies with loops are assumed for space graphs in mammals including man. While mathematically, the step from chains to networks may appear minor in nature, the presence of bifurcations (nodes with three or more links) requires an additional planning or choice mechanism between the possible transitions.

2.3 Place graph

The simplest type of graph representation of space is the place graph where each node represents a place in the world while the links are transversable connections between places. In Fig. 2a, we have assumed that these connections are directed, allowing to associate different travel instructions with each direction. Connections between places, if existing, need not be two-way, for example when it is possible to get from place A to place B immediately, while the reverse direction may require some deviation (Kuipers 1978, 2000, Gillner & Mallot 1998, Franz et al. 1998) [46, 47, 29, 22]. In animal and human behaviour, directionality of wayfinding knowledge is a wide-spread phenomenon. Knowing an outbound path does not necessarily imply that the way back is easily found. If the path is known, it may be quite different from the outbound route; for examples from ant navigation, see Kohler & Wehner (2005) [45], and Wehner et al. (2006) [85].

The notion of a "place", unanimous as it may seem, is not easily defined. Geometrically, a place may be identified with a point in some mathematical space, but in our context, places are memory items which have to be recognized based on some sort of landmark information. Empirically, knowledge of a place will always be described by an extended neighbourhood area of some anchor point, most notably the catchment area and the confusion area of a place. The confusion area is the set of points which cannot be distinguished from the goal location based on the cues represented in memory (see Hübner & Mallot, 2007) [39]. The catchment area is frequently studied in animal behaviour; it is a neighbourhood of the goal such that the animal is able to find the goal (or its confusion area) when released anywhere within the catchment area (Zeil et al., 2003) [91]. Catchment areas can be quite large, especially when distant landmarks are used for characterizing a place.

In neural network models of cognitive maps, place recognition is often modeled by "place fields" similar to the hippocampal place-fields known from rodent neurophysiology (O'Keefe et al., 1998, McNaughton et al., 2006) [61, 54]. In this case, the place is described by the neuron's firing field, i.e. a region of several tens of centimetres in diameter. This firing field is again a neighbourhood of the cell's 'preferred' place, together with a weighting function declining towards the margins of the region. Since the firing fields of different neurons overlap, the place cells constitute a population code of space (e.g., Arleo et al. 2004) [4]

The graph models used in architectural planning are also based on neighbourhoods of places, most notably the isovist or view-shed of a place (Benedikt, 1979) [8]. The isovist of a point comprises the set of all points that can be seen from one reference point. Points and their isovists can be connected in a visibility graph. In general, graphs based on catchment area and navigability will not be identical to graphs of isovists and mutual visibility, since navigable place transitions need not be visible from the starting point. Still, the isovist concept seems to capture some aspects

of human space perception, such as the appreciation of architectural spaces (Wiener et al., 2007) [87].

One problem shared by graph models form cognitive science and architectural planning is the selection of discrete points around which neighbourhoods are considered. In the animal literature, little is known about the selection of places to store in memory besides the primary goal locations such as nest and feeder; for one example from wood ants, see Durier et al. (2004) [17]. It seems likely, that simple optimality criteria apply.

2.4 View graph

While nodes of a place graph represent geometric points together with some neighbourhood, view graphs are based on the pose of the observer, i.e. on a combination of position and heading or body orientation. As in the place graph, this representation is not in terms of a coordinate vector, but rather in terms of a view of the environment associated with the heading direction and position as a vantage point.

At any one place, different views arise from different viewing directions. This is true even if the field of view spans the entire panorama, since the image coordinate system of the view will be centred around the heading direction. Indeed, despite their panoramic fields of view, wasps and other insects seem to use visual landmark information in a direction specific way, recognizing a place or panoramic view only if looked at from the direction in which it was encoded (Zeil et al. 1996) [92]. This might be due to the fact that view information is encoded jointly with geocentric compass information provided by the skylight polarization pattern (Rossel & Wehner 1982, Wehner & Müller 1985) [66, 86]. Evidence for directional encoding of positional information has also been presented for view-cells in the hippocampus of non-human primates, which seem to replace the mostly non-directional place-cells found in rodents (Rolls et al., 1998) [65]. In psychology, directional effects of place recognition have been presented by McNamara et al. (1984) [53], and Schweitzer et al. (1998) [71].

Views might also be associated with place connections, in the sense that a particular view of or from a place is obtained when approaching or leaving the place using a particular connection. As indicated in Fig. 2b, the nodes of the view graph correspond to links between places in a place graph. Schölkopf & Mallot (1995) [70] have presented a theory of cognitive map based entirely on such views. Connections between views indicate that during navigation, two views can be encountered in immediate temporal sequence. The view graph is a dual of the place graph in the sense that each place connection corresponds to a view and each complete bipartite subgraph of the view-graph corresponds to a place. Actions corresponding to view transitions can be coded in an egocentric way, such that coordinate transforms in the process of self-localization and route generation can be avoided. Learning a viewgraph amounts to accumulating schemata of the

form $(\text{view}_i, \text{action}_k, \text{view}_j)$ by adding nodes and links to the current graph. Exploration can be guided by the fact that the view-graph can be decomposed into complete bipartite subgraphs, each corresponding to a place.

One computational problem solved by the view-graph is that of snapshot alignment. Even if panoramic snapshots are taken to represent places, place recognition from unknown directions requires costly comparison operations such as calculating the complete circular cross-correlation function between the stored and current snapshots. Stürzl & Mallot (2006) [76] have suggested a coarse-to-fine approach to this problem, calculating the raw view alignment from the low spatial frequency components. The view-alignment problem shows that view-graph representations may be useful even if panoramic vision is present. In less than panoramic vision, it seems even more natural to base spatial memory on views. Indeed, Gaussier et al. (2002) [27] have used the view-graph in robot applications.

The notion of the view graph is also akin to the idea of the axial graph used by Hiller et al. (1993) [38] for describing road choices by drivers in central London. It should be noted, however, that urban environments are channeled in the sense that movement is restricted to a discrete, low number of directions at any one time. The view graph approach is not limited to this situation but may be used in open environments as well.

2.5 Hierarchies and regions

Evidence for the representation of regions in spatial longterm memory comes from a number of sources including human nearness judgments and path choices made on printed maps (Bailenson et al. 2000) [6] and optimal foraging in insects (Ohasi et al. 2006) [59]. In navigational studies with humans, Wiener & Mallot (2003) [88] and Wiener et al. (2004) [90] investigated way choices in environments offering simple way alternatives. Subjects were familiarized with an environment comprising a loop of six places, each place with an outward connection ending blindly in another place. Each place was marked with an object taken from one of three semantic groups. Places with objects from each group were grouped together forming semantic regions containing two adjacent places from the loop together with their dead end extensions. When asked to navigate from one place to the place directly opposing on the loop, subjects preferred the route initially staying in the semantic region of the start place and then immediately leading to the goal region. The alternative route would lead from the start region to the third, irrelevant region and only then into the goal region. The results clearly indicate that regional knowledge is represented in longterm memory and used in spatial planning. Wiener & Mallot (2003) [88] suggest that the underlying memory structure is a hierarchical graph with additional regional nodes connected to all place nodes comprising the region (Fig.2c). Thus, the hierarchical graph contains two types or nodes, places and regions, and three types of connections, place-to-place, place-to-region, and

region-to-region. Path planning uses the lowest hierarchy level at the starting point but may rely on coarser levels for more distant parts of the journey.

2.6 Ways and places

An alternative approach to space graphs that combines elements of the place and view graph together with aspects of hierarchies has been presented by Kuipers et al. (2003) [48], see Fig. 2d. In this "skeleton" approach, two types of nodes exist, one for places and another one for paths or ways. A place node is connected to a way node, if the way passes through the respective place. Overall, place and way nodes thus form a bipartite graph. Routes are not represented by place-to-place connections, but by connections of the type place-to-way and way-to-place. An interesting feature of this approach is that it can explain the emergence of "highways" which are used whenever possible, thus introducing a backbone or skeleton in the cognitive map. As in the space syntax approach mentioned above (Hiller et al. 1993) [38], the skeleton approach applies best to channeled spaces.

2.7 Landmarks as graph nodes

Throughout this section, we have considered nodes as spatial entities which have been occupied by the agent at some time or another. Places, poses, views, ways, and regions all are accessible to the agent and become nodes of the space graph only when the agent decides to store them as its current position. In contrast, landmarks need not be accessible spatial entities, either because they are unreachable or too far away (e.g., stars), or because they do not correspond to objects or places at all (e.g., view axes). We will discuss this notion of landmarks below (Section 4.1). However, some landmarks do correspond to accessible locations and may therefore be treated as nodes of a space graph. Indeed, this is the standard approach in robotics SLAM algorithms, where landmarks are localized in an occupancy grid and subsequently used for global metric embedding via triangulation. It should be noted, however, that this approach limits the use of landmarks, excluding possibly useful cues that do not lend themselves for localization.

3 Links and labels attached to links

Two graph nodes will be connected by a link, if it is known how to navigate from one node to the other. Links may be directional if only one travel direction can be carried out based on the knowledge stored in the graph. In the simplest case, the rule allowing navigation between two nodes may be the same rule used also for recognizing the place. In this case, all the agent needs to know is which neighbors can be found from each node and then start the search or homing

Figure 3: Places and landmarks. The photograph shows the view from a place called "four lakes view" near the town of Boppard, Germany. The four lakes are actually sections of the meandering river Rhein separated by mountain ridges. Thus, the name does not refer to landmark objects existing in the world (the "lakes"), but rather to the view obtained from the place. (photograph: F. König, http://de.wikipedia.org/wiki/Vierseenblicklift, GNU General Public Licence)

procedure. In the logic of snapshot-based memory, this means that a node v_1 is (directionally) connected to a node v_2 if and only if v_1 is contained in the catchment area of v_2 (Franz et al., 1998) [22].

Without further labeling, a link only contains the information that there is a way of navigating between the connected nodes. If this navigation requires special means or additional parameters, these have to be included in memory. One example of this is given by the "local vectors" in ant navigation which an ant can activate from memory to navigate from a recognized point to its nest (Bisch-Knaden & Wehner, 2003) [9]. The local vector contains information on distance and direction and is used just as a home-vector obtained from path integration. Unlike this vector, it is no longer subject to error accumulation once it has been correctly associated to that path. More generally, labels associated to links may be motor programs corresponding to the respective path segment. Just as the "local vector", such motor programs contain implicit distance and turning information which can be used in the construction of metric maps as described below. Also, labels attached to links may contain landmark information or guidances, such as in the procedure "follow the brick wall".

Allowing labels attached to links is one possibility of action representation. Alternatively, one might consider to attach these label to the nodes together with other labels discussed below. Indeed, in route memories, i.e., chain-like graphs, both possibilities are equivalent. In true maps (graphs containing nodes with three or more links), however, the following problem arises: Consider a label λ_i attached to a link from node v_{i1} to node v_{i2}. If this was to be attached to the start

link v_{i1}, it must take the form (λ_i, v_{i2}), since otherwise it would not be clear to which neighbour of v_{i1} the action specified by λ_i will lead. As a result, there will be two types of labels attached to node v_{i1}, one depending only on v_{i1} itself (these are the labels discussed in Section 4 below) and one depending on v_{i1} and another node from its neighbourhood. While this is possible, we consider it more parsimonious to attach label λ_i to the link itself, since this allows to treat chains and true graphs in the same way. In this scheme, each planning step from one node to the next is divided into two substeps, first the selection of the next node and second the retrieval of the required action (i.e., the label attached to the link to the next node).

4 Labels attached to graph nodes

4.1 Landmarks (local position information)

The most basic task in spatial behaviour is probably recognizing places. Even in simple search behaviour, the goal has to be recognized once it has been reached. In its simplest form, place recognition may be restricted to one or a few special places such as nest entries or feeding sites, but more complex spatial memories will contain larger numbers of known places. Place recognition has to rely on some sort of landmark information or distinct configuration of landmarks, i.e. sensor data characteristic of each place. Thus, the problem of place recognition is largely identical to the problem of landmark recognition. Let us define a landmark as a piece of sensory information, characteristic of a place, that is stored in memory and used in place recognition. Based on this definition, three questions about landmarks can be asked:

1. Depth of processing: What codes are generated from the sensory input to be stored in memory?
2. Landmark selection: Which parts of the input information are used to form landmark codes?
3. Landmark usage: How are landmarks used in spatial behaviour?

For the last question, it formally suffices to consider landmarks as characteristics of places, or, more generally, of the particular type of graph nodes considered in each model. The standard distinction between guidance (piloting) and direction (recognition-triggered response) made by O'Keefe and Nadel (1978) [62] and by Trullier et al. (1997) [80] can then be presented as two subsequent steps: The first step is pinning down the location of a place by landmark guidance, i.e. by moving such that a memorized view occurs on the navigator's retina. If this place (or state) is not the final destination, the next step will be a direction associated not with the landmark itself but with the graph link from the node recognized by means of the landmark to the next goal

node. As mentioned before, attaching the action label to the graph link rather than to the start node of the link allows to treat route and topological map knowledge in just the same way.

4.1.1 Depth of processing

The most general account of landmark information is the notion of "local position information", defined as the sum of all sensory inputs perceivable at a certain location (Trullier et al. 1997) [80]. Although this definition is not restricted to the visual modality, local views or snapshots will often be the most important type of local position information. In an extended environment, the local position information is a vector-valued function of position and pose (orientation of body and sensors) called the view manifold (Franz et al., 1998) [22]; the components of the vector are the sensor readings currently obtained. For the visual modality, the local position information is the image or view locally obtained. These views, parameterized with the agent's position and pose, thus form a manifold containing all visual landmark information available in the environment. More specific kinds of landmarks can be derived by applying various amounts of image processing (see Tab. 2). For robots, low-frequency Fourier components or the panoramic image can be used to bring snapshots into register for subsequent image comparison (Stürzl & Mallot, 2006) [76]. Honeybees have been shown to use raw image information ("snapshots") and find places by matching currently visible and remembered snapshots (Cartwright & Collett, 1982) [10]. In addition, Zhang et al. (1995) [93] demonstrated that honeybees are also able to match features defined by motion contrast or motion parallax. The usage of raw snapshot information in humans has recently been demonstrated in virtual reality experiments using a feature-less smooth colour cycle as a texture of the wall in a circular room. Subjects presented with the view visible from a given place in the room were able to approach this place from other locations (Gillner et al., *in press*) [30]. In addition, Christou and Bülthoff (1999) [13] have shown that landmark objects are recognized faster and more accurately if they are presented in a familiar rather than in a novel orientation, indicating that views rather than 3D objects are remembered in the navigation task.

In a mechanism called the geometric module (Cheng, 1986) [12], rats determine their position from local depth maps, i.e. the distances to the surrounding walls. This information is also derived from visual input but requires more elaborate processing. In rats, possible cues to determine the distance of walls include motion parallax, the elevation of the lower edge of the wall in the rat's visual image, and maybe some stereopsis. In humans, it was shown that young children (Hermer & Spelke, 1994) [36] as well as non-attentive adults (Hermer-Vazquez et al., 1999) [37] also use geometric information to find a hidden object in a room.

In everyday language, a landmark is an identified, nameable object or sight, whose recognition in an image requires a fully-fledged object recognition system. This type of information indeed increases navigation performance (Heft, 1981) [35]. The relation between landmark recognition

and object recognition in general is not entirely clear (e.g., Fig. 3). While many computational problems are identical, neural processing of the two recognition systems in humans shows an interesting dissociation (Janzen & van Tourennout, 2004) [42]: The recognition of attentively perceived objects is associated with activation in the right fusiform cortex, whereas objects at decision points are processed mainly in the parahippocampal region, indicating a difference in the neural processing underlying landmark- and object recognition.

The level with the largest depth of processing in the sketched landmark hierarchy consists in the assignment of names to the recognized landmarks. Nameability of landmarks may be more relevant in spatial language and direction giving, than in actual navigation.

Table 2: Depth of processing in landmark recognition.

input	image processing	output
view manifold	local recording	raw snapshot at discrete points
	edge detection	edge based snapshot
	depth estimate	depth signature (geometric module)
	object recognition	identified landmarks
	associating a text	named landmark

4.1.2 Landmark selection

As compared to the complete view manifold, or its processed versions, landmark information must be limited in two ways. First, only parts of each image should be stored as a memory. Second, landmarks should be recorded only at selected observer locations. Both selections will be guided by the following criteria:

1. **Salience:** Landmarks should be well recognizable. In verbal recall experiments, named objects are usually those standing out from the environment in size, shape, or functionality (Appelyard, 1970) [2].

2. **Relevance:** Landmarks should be remembered at places where navigational decisions are required. If adult subjects are asked to choose objects with potential landmark value they tend to name objects in the vicinity of road crossings (Allen et al., 1979) [1] where a decision has to be made. Note that the distinction of landmark and object recognition in (Janzen & van Tourennout, 2004) [42] was based on this criterion.

3. **Permanence:** The landmark and its position should be constant over time. In children, navigational errors have been shown to result from choosing as landmarks salient but non-permanent objects such as fancy cars, which may be gone when visiting their original location again (Cornell et al., 1994) [15]. Rat head direction cells follow the more distant of two independently moving landmarks even if they cover the same visual angle. A possible interpretation of this preference is that the more distant object will be larger and therefore less likely to move (Zugaro et al., 2001) [94].

4.2 Coordinates

In the graph approach, different types of metric representation can be considered. (i) In the most complete case, the full metric map, a node is introduced for each point in a sampled plane. These nodes are defined primarily by their coordinates and gain place or landmark information only after according evidence has been collected. A global coordinate system is needed to specifiy the coordinates of each point. The full metric map is used for example in robotic Simultaneous Localization and Mapping (SLAM) algorithms (Thrun, 1998) [77]. (ii) By "local metric information", we denote distance information between pairs of nodes, typically represented as labels of graph edges, and angular information associated to the central one of a triple of nodes (or pairs of edges). Graph nodes are introduced only for places, poses etc, which at some point have been occupied by the agent. Local metric information may be globally inconsistent and does not assume a global coordinate frame. (iii) "Metric embedding" is an extension of the local metric case using the same basic perceptual data, i.e. node recognition, node distances, and bearings. Metric embedding is intermediate in that local metric information is checked for consistency by local triangulation. Inconsistencies are removed by optimizing the local metric data or by assigning optimized coordinate values to each image point. In the first case, the representation remains coordinate-free, while in the second case, a global coordinate frame will be introduced.

Given the available experimental data summarized in Section 1.2, it appears that the human cognitive map uses local metric (type ii) and maybe some metric embedding (type iii). Behavioural evidence for complete metric maps (type i) seems to be missing. Metric perceptions in printed maps are probably closer to image processing than to navigation. An algorithm for recovering a metric embedding from local metric data has been presented by Hübner & Mallot (2007) [39].

As a caution remark, it should be noted that abandoning the idea of a global metric framework for the cognitive map makes it hard to interpret data on global orientation in maps. For example, Restat et al. (2004) [64] have shown that urban environments on slanted landscapes are learned together with the up-down-axis of the environment. One possibility might be to represent slope locally, i.e. to orient the snapshots of all nodes into the upward direction, say, as suggested by Jeffrey et al. (2006) [43] for rodent hippocampal place fields.

4.3 Functionality

In neuropsychology, the separation of a purely spatial component from other types of declarative longterm memory, especially the episodic memory[3], remains a matter of debate (Eichenbaum 1996) [19]. One point where spatial and more general memory types may meet is the issue of functionality, i.e. memories of places associated with actions that may be carried out at each place. With the addition of functionality labels, the graph structure discussed so far becomes a much more useful tool for the organization of behaviour, allowing to generate route plans from functional demands (see for example Arbib & Lieblich, 1977) [3].

In biology, the simplest function associated with a place seems to be the place's attractivity. A model for the formation of territories based on a pre-existing spatial map and learning of attractivity values has been suggested by Stamps & Krishnan (1999) [75]. An extension for simultaneous learning of places and place attractivities has been presented by Schmolke & Mallot (2002) [69]. Indeed, the cognitive representation of territories seems to be an evolutionary old trait which may even form the root of individual recognition. Bee & Gerhardt (2002) [7], for example, showed that territorial neighbours are tolerated by American bullfrogs as long as they are recognized in their proper territory. If two owners of adjacent territories change their calling location, they are recognized as intruders even so they would be tolerated when calling from their familiar sites. This result indicates that individual recognition may root in territoriality, which requires a memory of place function.

5 Multiple spatial memories

So far, we have considered the reference or longterm memory of space, of which the space graph appears to be a suitable model. Longterm memory is required for the recognition of places, as well as for route following and route planning. In addition to this reference memory, a behaving agent, both robot or animal, will need some sort of working memory as an interface between perception, reference memory, and behaviour. Working memory can be involved in spatial behaviour either in isolation or combined with long-term memory. We will therefore discuss theoretical approaches to working memory structure in this section.

5.1 Ego- and allocentric memories

One important distinction between different types of spatial memory is that between allocentric and egocentric memory frames. The best way of making this distinction seems to be in terms of each memory's behaviour upon movements of the observer. We call a representation of space or

[3] In neuropsychology, episodic memory is the memory for events which is considered a part of declarative memory. Since events always have a location, interactions between episodic and spatial memory may be expected.

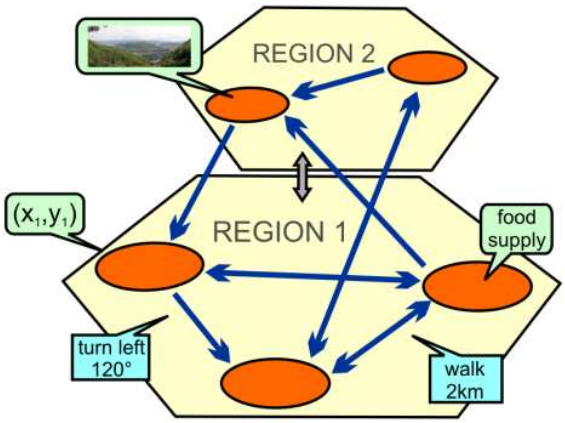

Figure 4: Summary of space-graph approach to spatial longterm memory. The nodes are assumed to be places, but the model can be generalized to view-graphs etc. Graph links are labeled by action rules providing local (implicit) metric information. For graph nodes, three types of labels are indicated, local position information, global metric coordinates, and functionality of places. Regions are symbolized by the hexagons, which may also be nodes of a hierarchical graph on a higher level.

objects "egocentric" if it changes in predictable ways as the observer moves, while the observer representation ("ego") remains unchanged. An egocentric representation could be a local map with the observer always at the origin, heading into the angular direction 0 degrees. If the observer ("ego") turns, the representation of outside objects will turn in the opposite direction, thus moving on circles about the origin. If "ego" translates, object representations will be translated in the opposite direction. The simplest type of an egocentric representation is the home-vector assumed in ant path integration (see Section 1.2). This home-vector is continuously updated during motion; it is thus a type of working memory where memory of preceeding steps is lost upon each update. Clearly, representations changing with each observer motion are not suitable for longterm memories which should be rather more stable.

The opposing term, "allocentric" representation, is used for representations of places or objects which do not change as the observer moves. In this case, the position of ego must be represented independently, as it will not stay in the centre of the map. Such representations do not necessarily need a coordinate frame with a well-defined "allo" as its centre. They may also be coordinate-free, e.g., based on distances and angles between objects or places. Still it will satisfy the requirement of stability with respect to observer motion. The term "allocentric" is thus missleading, as no centre needs to be defined. For a discussion of these terms, see also Klatzky (1998) [44]. Note that snapshots used as place-descriptors are allocentric in this definition even though they show

the place's environment in an egocentric view.

Due to interactions between ego- and allocentric mechanisms and representations, the distinction between both types may not always be clear-cut. One example is the metric embedding of local distance and angular information obtained from ego-centric path integration into an allocentric global map. Interactions between path integration and allocentric metric maps have been studied experimentally. Blindfolded human participants performing a path integration task have been found to show less uncertainty in familiar as compared to novel environment, indicating that prior knowledge of the environment is integrated with the ongoing path integration mechanism (Philbeck & O'Leary 2005) [63].

5.2 Working memory

5.2.1 Path integration

The best studied example for a spatial working memory is path integration, i.e. the ability of many animals to return to their nest in a straight line, even after a long, winding excursion. Unlike longterm memory, path integration does not require a permanent memory of the trail, but only a continuous update of a home-vector pointing from the agent's current position to the "home" position. In insects, this vector is generally assumed to be given in an egocentric reference frame (see Merkle et al. 2006 [58] for a review of models of path-integration in insects). In humans, it is generally assumed that solving path integration tasks involves more complex memories, but direct evidence for this assumption is scarce. The "encoding error model" (Fujita et al. 1993) [24], assumes that a sequence of egomotion steps is stored during the entire excursion and path integration is performed from these memories only when the return to home is being attempted. This model predicts that path integration should require longer processing time and be more prone to error for more complex excursions. However, this prediction could not be confirmed in behavioural experiments (Wiener & Mallot 2006) [89], arguing again for a continuous update of a home-vector.

In rodents, the recent finding of grid cells in the entorhinal cortex has fostered an alternative model of path integration, which may be called locally allocentric. Path integration takes place in a local metric map or chart, realized by a sheet of laterally connected neurons. As the agent moves, an activity peak representing the agent's position will be shifted on that map in the appropriate direction. When the margin of the map is reached, the activity peak will leave and reenter the map from the opposing side, i.e. the map is assumed to realize periodic boundary conditions or a toroidal topology. Thus, as the agent moves for long distances, the activity peak will reach the same neuron over and over again, resulting in a regular grid of firing fields of the map neurons (McNaughton et al. 2006) [54]. Indeed, the grid cells found in the rat entorhinal cortex by Hafting

et al. (2005) [31] show the expected properties.

5.2.2 Other working memory tasks

Other tasks related to path integration have been studied intensively in humans. In *spatial updating*, subjects are asked to predict the egocentric position of objects or places after performing some egomotion. This requires the same operations used for keeping track of the starting point in path integration. Indeed, in ants, it has been suggested that memories of home and feeding site are both build up by two separate vector memories, each being fed and updated with the same, instantaneous egomotion data (Collett & Collett 2000) [14]. Clearly, when trying to keep track of many objects, a local map model such as the locally allocentric grid-cell buffer seems rather more likely.

In *perspective taking* the subject is asked to imagine how a scene would appear when looked at from a different, not currently occupied view-point. This view-point is often marked by a person or an avatar in virtual reality studies. The computation is the same needed also in spatial updating, but it is not performed automatically, triggered by the actual movement of the observer, but under conscious control. Again, a local metric map might be a useful tool for achieving perspective taking performance. Perspective taking is often discussed in relation to social behaviour, considering the ability to infer what somebody else is seeing as a part of the Theory of Mind (Vogeley et al., 2004) [81]. However, direct evidence for this interpretation seems to be missing.

The prediction of views visible from viewpoints not currently occupied also plays a role in *route planning*, where decision points along a future route have to be foreseen. Clearly, route planning requires a tight interaction between reference and working memory. In the hierarchical graph model of Wiener & Mallot (2003) [88] a "focal map" is assumed as a working memory stage (see also Fig. 2c). This focal map contains the current observer position and its surrounding region at highest resolution, while other regions (higher levels of the hierarchical graph) are represented only be their regional nodes. As a result, planning will prefer routes passing through the lowest possible number of regions.

Evidence for route planning abilities in animals has been presented e.g. by Cramer & Gallistel (1997) [16] who had vervet monkeys choose between two routes leading to various amounts of food items. The monkey optimized their routes for overall food intake, taking into account at least two future steps.

In terms of the ego- vs. allocentric distinction, home-vectors and object representations spatially updated during bodily motion clearly classify as egocentric memories. In perspective taking and route planning, no actual motions are performed. The memories involved may still be called egocentric with respect to ego's imagined position.

5.3 Longterm memory

An important distinction in longterm memory pioneered by O'Keefe & Nadel (1978) [62] and Squire (1987) [74] is that between declarative ("locale", map-like) and non-declarative ("taxon", route-like) representations. In the graph approach, the distinction lies simply in the topology of the graph itself. If the graph is a chain, the memory will be of the route type, while for non-trivial graphs with bifurcations (i.e., graphs including nodes with more than two edges) the map type is reached. Declarative memory is thus not associated with individual nodes but is a property of the graph structure as a whole. The recall of declarative memory will always involve a planning stage retrieving a route or path through the graph. Indeed, in the graph view, the need for this additional planning stage marks the major distinction between route and map. The planning stage itself, however, is a type of working memory. Neural network implementations that have been proposed for the planning stage are based on standard algorithms for graph search; for examples see Schmajuk & Thiele (1992) [68] and Schölkopf & Mallot (1995) [70].

The distinction between map and route type memories has been confirmed by functional magnetic resonance imaging (fMRI) studies showing that planing of novel routes from map knowledge recruits hippocampal and other cortical areas while repetition of familiar routes recruits the caudate nucleus in the basal ganglia (Hartley et al. 2003) [32]. Similarly, in rats, stereotyped spatial behaviour is affected by pharmacological inhibition of the basal ganglia (Gengler et al. 2005) [28].

6 Conclusion: Hierarchical models of spatial memory and cognition

In this paper, we have presented a theory of spatial memory based on the graph approach, as summarized in Fig.4. The graph approach in turn was motivated by the cognitive needs of various navigators ranging from ants foraging from a central "home" place all the way to humans operating in complex environments where the spatial and social dimensions are closely intertwined.

The graph approach unifies various hierarchical theories of spatial cognition which are based on the linear sequence of places, routes, and maps. This linear hierarchy of places, routes, and maps has been discussed in a number of different contexts, including (i) logical and computational complexity (Kuipers 1978, 2000) [46, 47], (ii) the sequence of acquisition during learning and ontogeny (Siegel & White, 1975) [72], (iii) the modularity of neural systems (O'Keefe & Nadel, 1978) [62], and (iv) the complexity of behaviour in animals (Trullier et al., 1997, Mallot, 1999) [80, 50] and robots (Franz & Mallot, 2000) [21]. The existence of a logical hierarchy, however, does not necessarily imply that the same hierarchy should appear in the sequence of the acquisition of spatial knowledge or in the neural implementation. For example, spatial knowledge may be

acquired from a printed map and than be transferred to a route plan used for actual navigation. In this case, the acquisition sequence goes "backwards", from map to route. In general, the different types of hierarchies may differ substantially although significant dependencies may be expected. We argue that the graph approach will be useful to understand the relation between the different types of spatial hierarchies listed above.

Another advantage of the graph model lies in its applicability to navigational systems of different animal species and different complexity. For example, the spatial memories of insects like ants and honeybees can be modeled as graphs with star-topology, where the nest or hive is in the center and other known places are linked to the centre but not to each other. In humans, more complex memories have been demonstrated representing true networks of places and allowing the generation of novel routes from segments of known routes. The graph model can be scaled to these various types of spatial memory and therefore allows detailed comparisons.

The construction of cognitive systems in robotics can take advantage of biological theories of cognition in various ways. Besides copying partial solutions, the analysis of biological systems helps structuring the overall problem. In the design questions listed in Section 1.4, we have collected a number of issues where biological approaches seem to differ from robot approaches. The one field where we think interactions might be most promising is the representation of place functionalities, which can be used for joint planning of the spatial and non-spatial components of general action plans.

Acknowledgment. The work described in the paper was supported by the European Commission (FP6-2003-NEST-PATH Project "Wayfinding") and the Deutsche Forschungsgemeinschaft (SFB 550).

References

[1] Gary L. Allen, Kathleen C. Kirasic, Alexander W. Siegel, and James F. Herman. Developmental issues in cognitive mapping: The selection and utilization of environmental landmarks. *Child Development*, 50(4):1062–1070, 1979.

[2] D. Appleyard. Styles and methods of structuring a city. *Environment & Behavior*, 2:100 – 116, 1970.

[3] M. A. Arbib and Lieblich. Motivational learning of spatial behavior. In J. Metzler, editor, *Systems Neuroscience*, pages 221 – 239. Academic Press, New York, 1977.

[4] A. Arleo, F. Smeraldi, and W. Gerstner. Cognitive navigation based on nonuniform gabor space sampling, unsupervised growing network, and reinforcement learning. *IEEE Transactions of Neural Networks*, 15:639 – 652, 2004.

[5] A. Baddeley. *Working Memory*. Clarendon Press, Oxford, 1986.

[6] J. N. Bailenson, M. S. Shum, and D. H. Uttal. The initial segment strategy: a heurisitc for route selection. *Memory and Cognition*, 28:306 – 318, 2000.

[7] M. A. Bee and H. C. Gerhardt. Individual voice recognition in a territorial frog *(rana catesbeiana). Proceedings of the Royal Society (London) B*, 269:1443 – 1448, 2002.

[8] M. L. Benedikt. To take hold of space: Isovists and isovis fields. *Environment and Planning B*, 6:47 – 65, 1979.

[9] S. Bisch-Knaden and R. Wehner. Local vectors in desert ants: context-dependent landmark learning during outbound and homebound runs. *J. Comp. Physiol. A*, 189(3):181–187, 2003.

[10] B. A. Cartwright and T. S. Collett. How honey bees use landmarks to guide their return to a food source. *Nature*, 295:560 – 564, 1982.

[11] B. A. Cartwright and T. S. Collett. Landmark maps for honeybees. *Biological Cybernetics*, 57:85 – 93, 1987.

[12] K. Cheng. A purely geometric module in the rat's spatial representation. *Cognition*, 23:149 – 178, 1986.

[13] C. Christou and H. H. Bülthoff. View dependence in scene recognition after active learning. *Memory and Cognition*, 27:996–1007, 1999.

[14] M. Collett and T. Collett. How do insects use path integration for their navigation? *Biological Cybernetics*, 83:245–259, 2000.

[15] E. D. Cornell, C. D. Heth, and D. M. Alberts. Place recognition and way finding by children and adults. *Memory and Cognition*, 22:633 – 643, 1994.

[16] A. Cramer and C. Gallistel. Vervet monkeys as travelling salesmen. *Nature*, 387:464, 1997.

[17] Virginie Durier, Paul Graham, and Thomas S. Collett. Switching destinations: memory change in wood ants. *The Journal of Experimental Biology*, 207(14):2401–2408, 2004.

[18] Fred C. Dyer. Spatial memory and navigation by honeybees on the scale of the foraging range. *The Journal of Experimental Biology*, 199:147–154, 1996.

[19] H. Eichenbaum. Is the rodent hippocampus just for 'place'? *Current Opinion in Neurobiology*, 6:187 – 195, 1996.

[20] P. Foo, W. H. Warren, A. Duchon, and M. J. Tarr. Do humans integrate routes into a cognitive map? map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental psychology: Learning, Memory, and Cognition*, 31:195–215, 2005.

[21] M. O. Franz and H. A. Mallot. Biomimetic robot navigation. *Robotics and Autonomous Systems*, 30:133 – 153, 2000.

[22] M. O. Franz, B. Schölkopf, H. A. Mallot, and H. H. Bülthoff. Learning view graphs for robot navigation. *Autonomous Robots*, 5:111 – 125, 1998.

[23] M. O. Franz, B. Schölkopf, H. A. Mallot, and H. H. Bülthoff. Where did I take that snapshot? Scene–based homing by image matching. *Biological Cybernetics*, 79:191 – 202, 1998.

[24] N. Fujita, R. L. Klatzky, J. M. Loomis, and R. G. Golledge. The encoding-error model of path-way completion without vision. *Geographical Analysis*, 25:295 – 314, 1993.

[25] Shaun Gallagher. *How the Body Shapes the Mind*. Oxford University Press, USA, December 2005.

[26] C. R. Gallistel. *The organization of learning*. The MIT Press, Cambridge, MA, USA, 1990.

[27] P. Gaussier, A. Revel, J. P. Banquet, and V. Babeau. From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. *Biological Cybernetics*, 86:15–28, 2002.

[28] S. Gengler, H. A. Mallot, and C. Hölscher. Inactivation of the rat dorsal striatum impairs performance in spatial tasks and alters hippocampal theta in the freely moving rat. *Behavioral Brain Research*, 164:73 – 82, 2005.

[29] S. Gillner and H. A. Mallot. Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience*, 10:445 – 463, 1998.

[30] S. Gillner, A. M. Weiß, and H.A. Mallot. Visual place recognition and homing in the absence of feature-based landmark information. *Cognition*, in press.

[31] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436:801 – 806, 2005.

[32] T. Hartley, E. A. Maguire, H. J. Spiers, and N. Burgess. The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron*, 37:877 – 888, 2003.

[33] M. Hauser. *Wild Minds: What Animals Really Think*. Allen Lane The Penguin Press, London, 2000.

[34] M. D. Hauser, N. Chomsky, and W. T. Fitch. The faculty of language: what is it, eho has it, and how did it evolve? *Science*, 298:1569 – 1579, 2002.

[35] H. Heft. The role of environmental features in route-learning: Two exploratory studies of way-finding. *Environmental Psychology of Nonverbal Behavior*, 3(3):172–185, 1981.

[36] L. Hermer and E. S. Spelke. A geometric process for spatial reorientation in young children. *Nature*, 370:57 – 59, 1994.

[37] L. Hermer-Vazquez, E. S. Spelke, and A. S. Katsnelson. Sources of flexibility in human cognition: dual-task studies of space and language. *Cognitive psychology*, 39:3 – 36, 1999.

[38] B. Hillier, A. Penn, J. Hanson, T. Grajewski, and J. Xu. Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B*, 20:29–66, 1993.

[39] W. Hübner and H. A. Mallot. Metric embedding of view graphs. a vision and odometry-based approach to cognitive mapping. *Autonomous Robots*, 23:183–196, 2007.

[40] R. Hurlebaus, K. Basten, H.A. Mallot, and J.M. Wiener. Route learning strategies in a virtual cluttered environment. In C. Freksa et al., editor, *Spatial Cognition (SC) 2008*, number 5248 in Lecture Notes in Artificial Intelligence, pages 104–120. Springer, 2008.

[41] T. Ishikawa and D. R. Montello. Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, 52:93–129, 2006.

[42] G. Janzen and M van Turennout. Selective neural representation of objects relevant for navigation. *Nature Neuroscience*, 7:673 – 677, 2004.

[43] K. J. Jeffery, R. L. Anand, and M. I. Anderson. A role for terrain slope in orienting hippocampal place fields. *Experimental Brain Research*, 169:218–225, 2006.

[44] R. L. Klatzky. Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, and K. F. Wender, editors, *Spatial Cognition. An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, number 1404 in Lecture Notes in Artficial Intelligence, pages 1 – 17. Springer Verlag, Berlin, 1998.

[45] M. Kohler and R. Wehner. Idiosyncratic route-based memories in desert ants, *Melophorus bagoti*: How do they interact with path-integration vectors? *Neurobiol. Learn. Mem.*, 83:1–12, 2005.

[46] B. Kuipers. Modeling spatial knowledge. *Cognitive Science*, 2:129 – 153, 1978.

[47] B. Kuipers. The spatial semantic hierarchy. *Artificial Intelligence*, 119:191 – 233, 2000.

[48] B. J. Kuipers, D. G. Teucci, and B. J. Stankiewicz. The skeleton in the cognitive map. a computational and empirical exploration. *Environment and Behavior*, 35:81–106, 2003.

[49] George Lakoff. *Women, Fire, and Dangerous Things*. University Of Chicago Press, April 1987.

[50] H. A. Mallot. Spatial cognition: Behavioral competences, neural mechanisms, and evolutionary scaling. *Kognitionswissenschaft*, 8:40 – 48, 1999.

[51] H. A. Mallot and S. Gillner. Route navigation without place recognition: what is recognized in recognition–triggered responses? *Perception*, 29:43 – 55, 2000.

[52] T. P. McNamara and V. A. Diwadkar. Symmetry and asymmetriy of human spatial memory. *Cognitive Psychology*, 34:160–190, 1997.

[53] T. P. McNamara, R. Ratcliff, and G. McKoon. The mental representation of knowledge acquired from maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10:723 – 732, 1984.

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

[55] R. Menzel, K. Geiger, L. Chittka, J. Joerges, and J. Kunze. The knowledge base of bee navigation. *The Journal of Experimental Biology*, 199:141–146, 1996.

[56] R. Menzel, U. Greggers, A. Smith, S. Berger, R. Brandt, S. Brunke, G. Bundrock, S. Hülse, T. Plümpe, F. Schaupp, E. Schüttler, S. Stach, J. Stindt, N. Stollhoff, and S. Watzl. Honey bees navigate according to a map-like spatial memory. *PNAS*, 102:3040–3045, 2005.

[57] R. Menzel, R. J. De Marc, and U. Greggers. Spatial memory, navigation and dance behaviour in *Apis mellifera*. *J Comp Physiol A*, 192:889–903, 2006.

[58] T. Merkle, M. Rost, and W. Alt. Egocentric path integration models and their application to desert arthropods. *Journal of Theoretical Biology*, 240:385–399, 2006.

[59] K. Ohashi, J. D. Thomson, and D. D'Souza. Trapline foraging by bumble bees: IV. optimization of route geometry in the absence of competition. *Behavioural Ecology*, 2006. doi:10.1093/beheco/arl053.

[60] K. Ohashi, J. D. Thomson, and D. D'Souza. Trapline foraging by bumble bees. *Behavioural Ecology*, 18:1 – 11, 2007.

[61] J. O'Keefe, N. Burgess, J. G. Bennett, K. J. Jeffery, and E. A. Maguire. Place cells, navigational accuracy, and the human hippocampus. *Philosophical Transactions fo the Royal Society London B*, 353:1333–1340, 1998.

[62] J. O'Keefe and L. Nadel. *The hippocampus as a cognitive map.* Clarendon, Oxford, England, 1978.

[63] John W. Philbeck and Shannon O'Leary. Remembered landmarks enhance the precision of path integration. *Psicologica*, 26:724, 2005.

[64] J. Restat, S. D. Steck, H. F. Mochnatzki, and H. A. Mallot. Geographical slant facilitates navigation and orientation in virtual environments. *Perception*, 33:667 – 687, 2004.

[65] E. T. Rolls, A. Treves, R. G. Robertson, P. Georges-François, and S. Panzeri. Information about spatial view in an ensemble of primate hippocampal cells. *Journal of Neurophysiology*, 79:1797 – 1813, 1998.

[66] Samual Rossel and Rüdiger Wehner. The bee's map of the e-vector pattern in the sky. *PNAS*, 79:4451–4455, 1982.

[67] A. Schaub and H.-U. Schnitzler. Flight and ecolocation behaviour of three vespertilionid bat species while commuting on flyways. *Journal of Comparative Physiology A*, 193:1185–1194, 2007.

[68] N. A. Schmajuk and A. D. Thieme. Purposive behaviour and cognitive mapping: a neural network model. *Biological Cybernetics*, 67:165 – 174, 1992.

[69] A. Schmolke and H. A. Mallot. Territory formation in mobile robots. In R. K. Standish, M. A. Bedau, and H. A. Abbass, editors, *Artificial Life VIII*, pages 256 – 259, Cambridge, MA, 2002. The MIT Press.

[70] B. Schölkopf and H. A. Mallot. View–based cognitive mapping and path planning. *Adaptive Behavior*, 3:311 – 348, 1995.

[71] K. Schweizer, T. Herrmann, G. Janzen, and S. Katz. The route direction effect and its constraints. In C. Freksa, C. Habel, and K. F. Wender, editors, *Spatial cognition: an interdisciplinary approach to representing and processing spatial knowledge*, number 1404 in Lecture Notes in Computer Science, pages 19 – 38. Springer Verlag, Berlin, 1998.

[72] A. W. Siegel and S. H. White. The development of spatial representations of large-scale environments. In H. W. Reese, editor, *Advances in Child Development*, volume 10, pages 9–55. Academic Press, New York, 1975.

[73] E. S. Spelke and K. D. Kinzler. Core knowledge. *Developmental Science*, 10:89–96, 2007.

[74] L. R. Squire. *Memory and Brain.* Oxford University Press, New York, Oxford, 1987.

[75] J.A Stamps and V. V. Krishnan. A learning-based model of territory establishment. *Quarterly Review of Biology*, 74:291 – 318, 1999.

[76] W. Stürzl and H. A. Mallot. Efficient visual homing based on Fourier transformed panormaic images. *Robotics and Autonomous Systems*, 54:300–313, 2006.

[77] S. Thrun. Learning metric–topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 99:21 – 71, 1998.

[78] E. C. Tolman. *Purposive behavior in animals and men.* The Century Co., New York, 1932.

[79] E. C. Tolman. Cognitive maps in rats and man. *Psychological Review*, 55:189 – 208, 1948.

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

[81] K. Vogeley, M. May, A. Ritzl, P. Falkai, K. Zilles, and G. R. Fink. Neural correlates of first person perspective an one constituent of human self-consciousness. *Journal of Cognitive Neuroscience*, 16:815 – 827, 2004.

[82] K. von Frisch. *Tanzsprache und Orientierung der Bienen (engl. version: The Dance Language and Orientation in Bees(1967)).* Springer, Berlin, Heidelberg, New York, 1965.

[83] D. Waller and Y. Lippa. Landmarks as beacons and associative cues: their role in route learning. *Memory and Cognition*, 35:910–924, 2007.

[84] R. Wehner. Desert ant navigation: how miniature brains solve complex tasks. *JCPA*, 189:579 – 588, 2002.

[85] R. Wehner, M. Boyer, F. Loertscher, S. Sommer, and U. Menzi. Ant navigation: One-way routes rather than maps. *Current Biology*, 16:75–79, 2006.

[86] R. Wehner and M. Müller. Does interocular transfer occur in visual navigation by ants. *nature*, 315:228–229, 1985.

[87] J. M. Wiener, G. Franz, N. Rosmanith, A. Reichelt, H. A. Mallot, and H. H. Bülthoff. Isovist analysis captures properties of space relevant for locomotion and experience. *Perception*, 36:1066–1083, 2007.

[88] J. M. Wiener and H. A. Mallot. 'Fine-to-coarse' route planning and navigation in regionalized environments. *Spatial Cognition and Computation*, 3:331–358, 2003.

[89] J. M. Wiener and H. A. Mallot. Path complexity does not impair visual path integration. *Spatial Cognition and Computation*, 6:333–346, 2006.

[90] J. M. Wiener, A. Schnee, and H. A. Mallot. Use and interaction of navigation strategies in regionalized environments. *Journal of Environmental Psychology*, 24:475–493, 2004.

[91] J. Zeil, M. I. Hoffmann, and J. S. Chahl. Catchment areas of panoramic snapshots in outdoor scenes. *Journal of the Optical Society of America A*, 20:450 – 469, 2003.

[92] J Zeil, A Kelber, and R Voss. Structure and function of learning flights in bees and wasps. *J Experimental Biology*, 199:245–252, 1996.

[93] S.W. Zhang, M.V. Srinivasan, and T. Collett. Convergent processing in honeybee vision: Multiple channels for the recognition of shape. *Proc. Natl. Acad. Sci.*, 92:3029–3031, 1995.

[94] M. B. Zugaro, A. Berthoz, and S. I. Wiener. Background, but not foreground, spatial cues are taken as references for head direction responses by rat anterodorsal thalamus neurons. *The Journal of Neuroscience, 2001, 21:RC154:1-5*, 21(RC154):1–5, 2001.