The Network of Reference Frames Theory: A Synthesis of Graphs and Cognitive Maps

Tobias Meilinger

Max-Planck-Institute for Biological Cybernetics
Spemannstr. 44, 72076 Tübingen, Germany
tobias.meilinger@tuebingen.mpg.de

Abstract. The network of reference frames theory explains the orientation behavior of human and non-human animals in directly experienced environmental spaces, such as buildings or towns. This includes self-localization, route and survey navigation. It is a synthesis of graph representations and cognitive maps, and solves the problems associated with explaining orientation behavior based either on graphs, maps or both of them in parallel. Additionally, the theory points out the unique role of vista spaces and asymmetries in spatial memory. New predictions are derived from the theory, one of which has been tested recently.

Keywords: graph; cognitive map; spatial memory; reference frame; route knowledge; survey knowledge; self-localization; environmental space.

1 Introduction

Orientation in space is fundamental for all humans and the majority of other animals. Accomplishing goals frequently requires moving through environmental spaces such as forests, houses, or cities [26]. How do navigators accomplish this? How do they represent the environment they traveled? Which processes operate on these representations in order to reach distant destinations or to self-localize when lost? Various theories have been proposed to explain these questions. Regarding the underlying representation these theories can be roughly classified into two groups which are called here graph representations and cognitive maps. In the following paper, I will explain graph representations and cognitive maps. I will also highlight how graph representations and cognitive maps fail to properly explain orientation behaviour. As a solution I will introduce the network of reference frames theory and discuss it with respect to other theories and further empirical results.

1.1 Graph Representations and Cognitive Maps

In graphs the environment is represented as multiple interconnected units (e.g., [4], [19], [20], [45], [48]; see Fig. 1). A node within such a graph, for example, represents a location in space or a specific sensory input encountered, such as a view. An edge within a graph typically represents the action necessary to reach the adjacent node. Graphs are particularly suitable for explaining navigating and communicating routes (i.e., a sequence of actions at locations or views which allows navigators at a location

C. Freksa et al. (Eds.): Spatial Cognition VI, LNAI 5248, pp. 344–360, 2008.

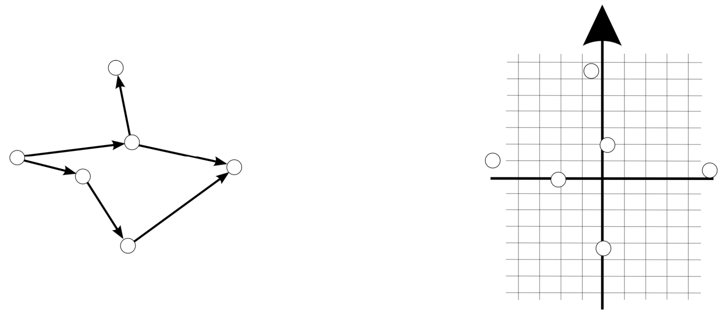

Fig. 1. Visualizations of a graph representation where an environment is represented as multiple interconnected units (left) and a cognitive map where an environment is represented within one reference frame (right)

A to reach B without necessarily knowing where exactly B is relative to A). This could be, for example,, turn right at the church, then turn left at the next intersection etc., The knowledge expressed in these sequences is called route knowledge.

A cognitive map, on the other hand, assumes that the environment is represented within one single metric frame of reference, (i.e., all locations within the environment can be expressed by coordinates of one single coordinate system; see Fig. 1; [2], [7], [30]; cf., [28], [33]).[1] A cognitive map has to be constructed from several different pieces of information encountered during navigation. The case of learning a cognitive map from a physical map which provides the information already within one frame of reference is not considered here. A cognitive map is especially suited to provide direct spatial relations between two locations, without necessarily knowing how to get there, for example, the station is 300 Meters to the east from my current location. This type of knowledge is known as survey knowledge. Survey knowledge is necessary for tasks such as shortcutting or pointing to distant locations.

1.2 Problems with Graph Representations and Cognitive Maps

Graph representations and cognitive maps are especially suited to represent route and survey knowledge, respectively. The other side of the coin is, however, that they also have their specific limitations. These will now be described in detail.

Graph representations (1) do not represent survey knowledge, (2) often ignore metric relations given in perception, and (3) often assume actions are sufficient to explain route knowledge. The main limitation of graph representations is that there is no survey knowledge expressed at all. Using a graph representation, navigators know how to reach a location and have the ability to choose between different routes. Graph representations, however, do not give navigators any cue as to where their goal is

[1] Often the term cognitive map is used for the sum of all spatial representations. Contrary to that, cognitive map is understood here as a specific spatial representation, namely storing spatial information within one reference frame. A reference frame here is not understood as the general notion of representing something relative to ones' own body (= egocentric) vs. relative to other objects (= allocentric), but a reference frame is considered as one single coordinate system (cf. [15]). Nevertheless, a reference frame can be egocentric or allocentric.

located in terms of direction or distance. This problem originates from the fact that graph representations normally do not represent metric knowledge at all. This is despite the fact that not only human navigators are provided with at least rough distance estimates especially by their visual system and by proprioceptive cues during locomotion. Some graph models ignore this already available information and instead assume that a navigator stores raw or only barely processed sensory data ([4], [20]). As a final point, actions themselves ([19], [20], [45]) can not be sufficient to explain route knowledge. Rats can swim a route learned by walking [18]. Cats can walk a route learned while being passively carried along a route [10]. We can cycle a path learned by walking. Even for route knowledge the edge of a graph representing how to get from one node to the next has to be more abstract than a specific action. However, not only graph representations are limited.

Cognitive maps (1) have problems in explaining self-localization and route knowledge. (2) There is a surprising lack of evidence that proves non-human animals have cognitive maps at all. (3) Human survey navigation is not always consistent with a cognitive map, and (4), cognitive maps are necessarily limited in size. Self-localizing based exclusively on a cognitive map can only take the geometric relations into account that are displayed there, (e.g., the form of a place). The visual appearance of landmarks is almost impossible to represent within a cognitive map itself. This information has to be represented separately and somehow linked to a location within the cognitive map. This is probably one reason why simultaneously constructing a map while staying located within this map (SLAM) is considered a complicated problem in robotics [42]. Similarly, planning a route based on a cognitive map alone is also not trivial, as possible routes have to be identified first [16]. Another issue is that cognitive maps seem to be limited to human navigation. If animals had cognitive maps, they would easily be able to take novel shortcuts, (i.e., directly approach a goal via a novel path without using updating or landmarks visible from both locations). However, the few observations arguing for novel shortcuts in insects and mammals have been criticized because they do not exclude alternative explanations and could not be replicated in better controlled experiments [1]. For example, in the famous experiment by Tolman, Ritchie and Khalish [43] rat's shortcutting behavior can be explained by assuming they directly approached the only available light source within the room. Although the discussion whether non-human animals are able to make novel shortcuts has yet to be settled, such shortcutting behavior should be fairly common if orientation was based on a cognitive map. This is clearly not the case. Similarly, a human shortcutting experiment within an "impossible" virtual environment casts doubt upon a cognitive map as the basis for such survey navigation [34]. In this experiment unnoticeable portals within the virtual environment teleported participants to another location within the environment. They could, therefore, not construct a consistent two-dimensional map of this environment. Still, participants were able to shortcut quite accurately. The last shortcoming of cognitive maps is that we have to use many of them anyway. We surely do not have one and the same cognitive map (reference frame) to represent the house we grew up in, New York and the Eiffel Tower. At one point, we have to use multiple cognitive maps and (probably) represent relations between them.

Graph representations and cognitive maps have specific advantages and limitations. Graphs are good for representing route knowledge. However, they do not explain survey knowledge. Contrary to that, cognitive maps are straight forward representations of survey knowledge. They are, however, not well suited for self-localization and route knowledge and fail to explain some human and non-human orientation behavior. As a solution to these limitations, often both representations are assumed in parallel to best account for the different behaviors observed ([2], [4], [30], [45]; see also [12], [28], [33]). However, assuming two representations in parallel also poses difficulties. First, the last three arguments against cognitive maps also argue against theories which assume both graphs and cognitive maps. In addition, according to "Occam's razor" (law of parsimony), one simple representation is preferable to multiple representations when explaining behavior. Multiple representations of one environment also raise the question of how these representations are connected. A house for example, can be represented within a graph and a cognitive map. The house-representation in the map should refer to the corresponding house representation within the graph and not to a representation of another house. First, this correspondence has to be specified somehow, for example, via an association which results in even more information to be represented. Second, it is a non-trivial problem to keep the correspondences free of error. A theory has to state how this is accomplished.

In conclusion, neither a graph representation, nor a cognitive map alone is sufficient to convincingly explain orientation behavior in humans and non-human animals. Both representations together also pose tremendous difficulties. As a solution to these problems, I would like to propose the network of reference frames theory which combines graphs and cognitive maps within one representation. This theory described in Chapter 2 avoids the problems which were already mentioned. Together with processes operating on this representation, it explains self-localization, route navigation and survey navigation. Furthermore, this theory can also explain other effects which have not yet been pointed out. This will be described in Chapter 3 where it will also be compared to other theories.

2 The Network of Reference Frames Theory

In this chapter, I will describe the network of reference frames theory in terms of the representations and the processes acting on those, and how these are used for different tasks, such as navigation, survey knowledge, etc.

2.1 Representation

The network of reference frames theory describes the memory representation acquired by human and non-human animals when locomoting through environmental spaces such as the country side, buildings, or cities. It also describes how this representation is used for self-localization, route and survey navigation. The theory is a fusion between graph representations and cognitive maps (cf., Fig.2). It assumes that the environment is encoded in multiple interconnected reference frames. Each reference frame can be described as a coordinate system with a specific orientation. These reference frames form a network or graph. A node within this network is a reference frame referring to a single vista space. Vista spaces surround the navigator and can be perceived from

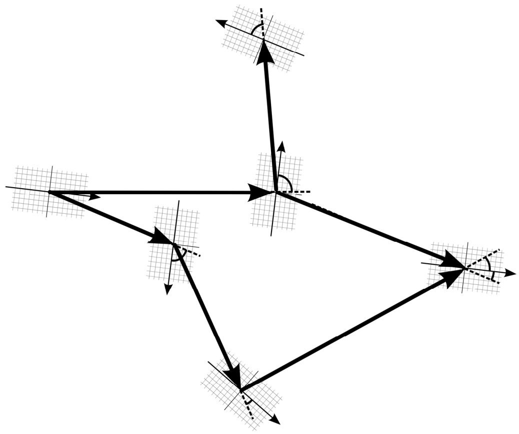

Fig. 2. A visualization of the network of reference frame theory. Reference frames correspond to single vista spaces. They are connected via perspective shifts which specify the translation and rotation necessary to get from one reference frame to the next one.

one point of view, for example, a room, a street or even a valley [26].[2] This means that the basic unit in the network is always the reference frame of a vista space. Within this vista space reference frame, the location of objects and the surrounding geometry are specified. The edges in the network define the so called perspective shift necessary to move from one reference frame to the next. Such a perspective shift consists of both a translation and a rotation component, for example, moving forward 150 meters and then turning right 90°. Perspective shifts all point to another reference frame,[3] they may differ in precision and the association strength with which they connect the two reference frames. The more familiar a navigator is with an environment, the more precise the perspective shifts will become and the more strongly the perspective shift will connect two reference frames.

The network of vista space reference frames connected via perspective shifts is stored in long-term memory. Several processes shape or operate on this memory. These processes are encoding, reorientation by recognition, route navigation, and survey navigation. In the following they will be described in detail (for a summary see Table 1).

2.2 Encoding

First Time Encounter. Encoding describes the process of constructing a representation of an environmental space through initial and continued contact. It is assumed

[2] Vista spaces extend to the back of a navigator (although nothing might be represented there). While other senses such as audition or information from self motion may be used to construct a representation of a vista space, the main source to do so will be vision.

[3] Humans are able to imagine how a perceived or a remembered vista space looks like from a different perspective. Such an imaginary shift in perspective within a vista space is not what is called perspective shift in the network of reference frames theory. Here a perspective shift, first, is stored in memory and is not imagined online, and second, a perspective shift always connects two vista spaces and does not occur within one vista space.

that encoding happens automatically. When navigating through an environmental space for the first time, we perceive vista spaces within the environmental space. This perceived vista space corresponds to a reference frame. The orientation of that reference frame is either determined by the view from which the vista space was experienced in the first place (cf., [20], [46]) or it is determined by the salient geometry of that vista space ([28], [33]). In daily life, these two directions usually coincide. For example, when entering a street or a house, our first view of the street or house is usually aligned with the geometry of the surrounding walls. Accessing such a reference frame will be easier and lead to an improved performance when one is aligned with the orientation of this reference frame, (e.g., looking down the street), than when not aligned, (e.g., facing a house in the street). Within this reference frame, the geometry of the enclosure is encoded (e.g., walls, hedges, houses or large objects). In addition to the geometry, locations of objects, such as landmarks, can be located within such a reference frame of a vista space.

After encoding an individual reference frame, a navigator moves on and encodes other reference frames corresponding to other vista spaces. These vista spaces do not necessarily have to be adjacent. A perspective shift will connect the two vista space reference frames, (i.e., the translations and rotations necessary to get from the first reference frame to the second). This perspective shift can be derived (1) from the visual scene itself, (2) from updating during navigating between the two vista spaces, and (3) from global landmarks visible from both vista spaces.

Deriving the perspective shift from the visual scene can be shown in an example such as standing in the corridor of a house and watching the kitchen door. The kitchen

Table 1. Summary of the representation and processes assumed in the network of reference frames theory

Representation
Network (graph) consisting of nodes connected by edges (see Fig. 2)
Node: a reference frame with an orientation specifying locations and orientations within a vista space; within this reference frame, objects and the geometric layout are encoded
Edge: perspective shift, i.e., translation and rotation necessary to move to the next reference frame; perspective shifts point to the next reference frame and differ in precision and association strength.
Processes
Encoding: first time experience or the geometry of a vista space define the orientation of a new reference frame; the visual scene itself, updating, or global landmarks can provide the perspective shift to the next vista space reference frame; familiarity increases the accuracy of the perspective shifts and the association strength of these connections.
Self-localization by recognition: recognizing a vista space by the geometry or landmarks it contains provides location and orientation within this vista space and the current node/reference frame within the network
Route navigation by activation spread: an activation spread mechanism provides a route from the current location to the goal; during wayfinding, reference frames on the route are pre-activated and, therefore, recognized more easily; recently visited reference frames are deactivated
Survey navigation by imagination: imagining connected vista spaces not visible step-by-step within the current reference frame; allows retrieving direction and straight line distance to distant locations; this can be used for shortcutting or pointing.

door provides us with the information of where (translational component) and in which orientation (rotational component) the kitchen is located with respect to the reference frame of the corridor. Extracting the perspective shift from the visual scene itself, however, only works for adjacent vista spaces with a visible connection.

For non-adjacent vista spaces, updating can provide the perspective shift. In doing so, one's location and orientation within the current reference frame is updated while moving away from its origin, (i.e., navigators track their location and orientation relative to the latest encoded vista space). When encoding a new reference frame, the updated distance and orientation within the former reference frame provides the necessary perspective shift to get from the first reference frame to the next. In that sense, updating can provide the "glue" connecting locations in an environmental space (cf., [17]). Updating can also work as a lifeline saving navigators from getting lost. As long as navigators update the last reference frame visited, they are able to return to the origin of the last encoded reference frame, (i.e., they are oriented).

A third possibility to get a perspective shift when already located in the second reference frame is by self-localizing with respect to a global landmark also visible from the first vista space reference frame, for example, a tower or a mountain top. Self-localizing provides a navigator with the position and orientation with respect to the reference frame in which the global landmark was first experienced. This is the perspective shift necessary to get from the first reference frame to the second one.

Repeated Visits. Re-visiting an environmental space can add new perspective shifts to the network and will increase the precision and association strength of existing perspective shifts (for the later see 2.4). Walking a new route to a familiar goal will form a new chain of reference frames and perspective shifts connecting the start and goal. That way, formerly unconnected areas, such as city districts, can be connected. When walking a known route in reverse direction, the theory assumes that new perspective shifts are encoded in a backward direction. Then two reference frames A and B are connected with two perspective shifts, one pointing from A to B and the other one pointing from B to A. In principle, inverting one perspective shift would be sufficient to get the opposite perspective shift. However, such an inversion process is assumed to be error-prone and costly therefore it is usually not applied.

When navigating an existing perspective shift along its orientation repeatedly no new perspective shift is encoded, but the existing perspective shift becomes more precise. This increase in precision corresponds to a shift from route knowledge to more precise survey knowledge. The precision of survey knowledge is directly dependent upon the precision of the perspective shift (for a similar model for updating see [6]). For many people, perspective shifts will be imprecise after the first visit, and therefore, highly insufficient, (e.g., for pointing to distant destinations). However, they still accurately represent route knowledge, (i.e., indicate which reference frame is connected with which other reference frame). When the perspective shifts become more precise after repeated visits, survey knowledge will also become more precise (cf., [25]; see 2.5). This corresponds with the original claim that route knowledge usually develops earlier than survey knowledge (e.g., [36]). However, survey knowledge does not have to develop at all (e.g., [24]) or can in principle also be observed after just a few learning trials (e.g., [27]). Correspondingly, the perspective shifts may be precise enough for pointing or other survey knowledge tasks after little experience or they may remain imprecise even after an extended experience. Here, large differences between individuals due to the sense of direction can be expected (cf., [9], [35]). Updating global orientation while navigating an environmental space will result in more precise perspective shifts, and therefore, improve survey knowledge. It follows that people with a good sense of direction will also acquire precise survey knowledge quicker. Similarly, environments which ease such updating will lead to more precise perspective shifts and improve survey knowledge accordingly. This facilitation can be gained, for example, by uniform slant, distant landmarks, or a grid city, which all have been shown to enhance orientation performance (e.g., [25], [32]).

2.3 Self-localization by Recognition

When someone gets lost within a familiar environmental space, the principal mode of reorientation will be by recognizing a single vista space within this environment (for self-localizing by the structure of environmental spaces see [21], [38]). A vista space can be recognized by its geometry or by salient landmarks located within (cf. [3]). First, recognizing a vista space provides navigators, with their location and their orientation within this vista space. Second, recognizing a vista space provides navigators with their location within the network, (i.e., in which node or vista space reference frame they are located). Their position in terms of direction and distance with respect to currently hidden locations in the environmental space however, has to be inferred from memory. This will be explained in the section on survey navigation by imagination further below.

2.4 Route Navigation by Activation Spread

Route navigation means selecting and traveling a route from the current location to a goal. The network of reference frames theory assumes an activation spread mechanism to explain route selection which was proposed by Chown et al. [4] as well as Trullier et al. [45]. Within the network, activation from the current reference frame (current node) spreads along the perspective shifts (edges) connecting the various reference frames (nodes). If the activation reaches the goal node, the route transferring the activation will be selected, (i.e., a chain of reference frames connected with perspective shifts). Here, the association strength of perspective shifts is important. The association strength is higher for the most navigated perspective shifts. Activation will be spread faster along those edges that are higher in association strength. If several possible routes are encoded within the network, the route that spreads the activation fastest will be selected for navigation. This route must not necessarily be the shortest route or the route with the least number of nodes. As the activation propagates easier via highly associated edges, such familiar routes will be selected with higher probability.

During navigation, the perspective shift provides navigators with information about where to move next, (i.e., perform the perspective shift). If the perspective shift is rather imprecise, navigators will only have an indicated direction in which to move. Moving in this direction, they will eventually be able to recognize another vista space reference frame. By updating the last reference frame visited, it will prevent navigators from getting lost. Pre-activating reference frames to come and de-activating already visited reference frames will facilitate recognition. When successfully navigating a known route, its perspective shifts will become more accurate and their association strengths will increase, making it more probable that the route will be selected again.

The described process is probably sufficient to explain most non-human route navigation. It is also plausible that such a process is inherited in humans and applied for example, when navigating familiar environments without paying much attention. However, humans can certainly override this process and select routes by other means.

2.5 Survey Navigation by Imagination

Survey knowledge tasks such a pointing or shortcutting require that relevant locations are represented within one frame of reference, (e.g., the current location and the goal destination). The network of reference frames theory assumes that this integration within one frame of reference occurs online within working memory. This is only done when necessary and only for the respective area. For example, when pointing to a specific destination, only the area from the current location to the destination is represented. In this framework, the integration within one frame of reference happens during the retrieval of information and not during encoding or elaboration, as with a cognitive map. The common reference frame is available only temporarily in working memory and is not constantly represented in long term memory. The integration itself is done by imagining distant locations as if the visibility barriers of the current vista space were transparent. The current vista space can be the one physically surrounding the navigator or another vista space that is imagined. From the current vista space's reference frame, a perspective shift provides the direction and orientation of the connected reference frame. With this information, the navigator imagines the next vista space within the current frame of reference, (i.e., this location is imagined in terms of direction and distance from the current vista space). This way, the second vista space is included in the current reference frame. Now, a third vista space can be included using the perspective shift connecting the second and the third vista space reference frames. That way, every location known in the surrounding environmental space can be imagined. Now, the navigator can point to this distant location, determine the straight line distance, and try to find a shortcut.

3 The Network of Reference Frames Theory in the Theoretical and the Empirical Context

3.1 The Network of Reference Frames Theory Compared to Graph Representations and Cognitive Maps

The network of reference frames theory is a fusion between graph representations and cognitive maps. Multiple reference frames or cognitive maps are connected with each other within a graph structure. As in graph representations, the basic structure is a network or graph. However, in contrast to most existing graph models ([4], [19], [20], [45], [48]), metric information is included within this graph. This is done for the nodes, which consist of reference frames, as well as for the edges, (i.e., the perspective shifts, which represent translations and turns). Such a representation avoids the problems associated with the mentioned graph representations (see 1.2): (1) Most importantly, it can explain survey knowledge, as metric relations are represented contrary to other graph models. (2) Representing metric relations also uses information provided by perception. Depth vision and other processes allow us to perceive the spatial structure of a scene. This information is stored and not discarded like in other graph models. (3) Perspective shifts represent abstract relations that can be used to guide walking, cycling, driving, etc. No problem of generalizing from one represented action to another action occurs as in other graph representations.

The network of reference frames theory also avoids problems from the cognitive map (see 1.2): (1) It can explain self-localization and route navigation in a straight forward manner which is difficult for cognitive maps. (2) An environmental space is not encoded within one reference frame as with a cognitive map. The representation, therefore, does not have to be consistent globally. So, contrary to cognitive maps, short cutting is also possible when navigating "impossible" virtual environments [34]. (3) The lack of clear evidence for survey navigation in non-human mammals and insects can be easily explained. According to the network of reference frames theory, these animals are not capable of imagining anything or they do not do so for survey purposes. However, survey navigation relies on the same representation as self-localization and route navigation. Only the additional process of imagining operates on this representation. This process might have even evolved for completely different purposes than navigation. Contrary to that, cognitive map theory has to assume that an additional representation, (i.e., a cognitive map), evolved only in humans specifically for orientation. These are much stronger assumptions. (4) Imagining distant destinations within working memory involves a lot of computation. Survey tasks are, therefore, rigorous and error prone which probably most people can confirm. In contrast, this daily life observation is not plausible with a cognitive map. Deriving the direction to distant locations from a cognitive map is rather straight forward and should not be more rigorous than, for example, route navigation. [4]

The network of reference frames theory also has advantages compared to assuming both a graph and a cognitive map in parallel (see 1.2):[5] Here survey navigation is again explained by the cognitive map part. This does not avoid the last three problems mentioned in the last paragraph.[6] In addition, the network of reference frames theory makes fewer assumptions. On a rough scale, it only assumes one representation, the

[4] Alternatively to simply read out survey relations from a cognitive map, mental travel has been proposed as an alternative process [2]. Mental travel can be considered as being more effortful and is, therefore, much more plausible. For the network of reference frames theory continuous mental travel in the area of an encoded vista space can be imagined. Between non-adjacent vista spaces, this should be rather difficult.

[5] Some theories assuming both a network representation and a global cognitive map are skeptical regarding the necessity and the evidence for such a cognitive map ([16], [31]).

[6] In his theory Poucet [31] assumes a network layer with pairwise metric relations between places. This representation can be used to compute shortcuts and avoids the problems mentioned with cognitive maps. However, Poucet also proposes a global integration within a cognitive map, leading again to the mentioned problems. In addition, it is unclear which of the two metric representations determine survey navigation.

combination of graphs and maps assume two representations. More specifically, graphs and maps need to connect corresponding elements, for example elements which represent the same house. These connections are extra and are potentially more error prone. A last problem with cognitive maps already mentioned is that we must have multiple cognitive maps anyway, because we cannot represent the whole world within one single cognitive map. As we do use reference frames to represent spatial locations, the question is, what spatial area do such reference frames encode usually? Here, it is proposed that this basic unit consists of a vista space.

3.2 Vista Space Reference Frames as the Basic Unit in the Representation of Environmental Spaces

Representing a space in multiple interconnected units, works with units of different size. Using large units such as towns, results in large packages of information which might be difficult to process as a whole. On the other hand, smaller units such as individual objects, result in an exponential increase in relations between the units which have to be represented. Many experiments show that humans are able to represent vista spaces within one frame of reference (e.g., [12], [28], [33]). So the main question is whether navigators use vista spaces or whether they also use larger units, (e.g., buildings or city districts), to represent locations within one reference frame. Updating experiments indicate that a surrounding room is always updated during blindfolded rotations. This is not necessarily the case for the whole surrounding campus suggesting that the relevant unit is smaller than a campus [47].

The network of reference fames theory predicts that there are no common reference frames for units larger than vista spaces. Other theories on spatial orientation in robots [50] and rodents [44] also rely on the visible area as the basic element.[7] Several arguments support vista spaces as the basic unit in spatial orientation. (1) Vista spaces are the largest unit provided directly by visual perception, and (2) they are directly relevant for navigation. (3) Visibility is correlated with wayfinding performance. (4) Hippocampal place cells are likely related to vista spaces, and (5) our own experiments show that participants encode a simple environmental space not within one reference frame, but use multiple reference frames in the orientation predicted by the network of reference frames theory.

Vista spaces can be experienced from only one point of view. In order to represent environmental spaces, such as buildings and cities, we have to move around (the case of learning from paper maps is not considered here). When encoding units larger than vista spaces, several percepts have to be integrated. Such integration is not done spontaneously [8]. Vista spaces are also the most relevant unit for navigation. Route decisions have to be taken within a vista space. When lost, self-localization is usually accomplished by recognizing the geometry or landmarks within a specific vista space

[7] In Yeap's theory [50] all vista spaces are directly adjacent to each other and are connected via exits. Survey relations computed from that representation are, therefore, correct when the form of individual vista spaces are correct. In the network of reference frames theory the preciseness of survey relations depends of the preciseness of the perspective shifts. In addition, Yeap assumes a hierarchical structuring on top of the basic vista space level. Touretzky and Redish [44] do not tell anything about environmental spaces. They also assume multiple, simultaneously active reference frames represent one vista space.

(e.g., [3]). Short cutting is difficult, because it encompasses more than just one vista space. In contrast, selecting the direct path to a necessarily visible location within a vista space is trivial. Visibility is also correlated with behavior. More vista spaces, (i.e., corridors on a route), lead to larger errors in Euclidean distance estimation [41]. Learning a virtual environmental space is easier with a full view down a corridor than when visual access is restricted to a short distance, which results in more vista spaces that need be encoded [38]. Place cells in human and rodent hippocampus seem to represent a location in a vista space ([5], [30]). Place cells fire every time a navigator crosses a specific area independent of head orientation. This area is relative to the surrounding boundaries of a vista space and is adjusted when changing the overall size or shape of the vista space [29]. One and the same place cell can be active in different vista spaces, and can therefore, not encode one specific location in an environmental space [37]. In conclusion, a set of place cells is a possible neuronal representation of locations within one frame of reference. This frame is likely to be limited to a vista space.

In addition to arguments from the literature, we recently tested the prediction from the network reference frames theory concerning the importance of vista space reference frames [23]. This prediction incorporated, first, that a vista space is the largest unit encoded within one single reference frame, and second, that the orientation of such a vista space reference frame is important, (i.e., that navigators perform better when they are aligned with that orientation). Participants learned a simple immersive virtual environmental space consisting of seven corridors by walking in one direction. In the testing phase, they were teleported to different locations in the environment and were asked to self-localize and then point towards previously learned targets. As predicted by the network of reference frames theory, participants performed better when oriented in the direction in which they originally learned each corridor, (i.e., when they were aligned with an encoded vista space reference frame). If the whole environment was encoded within one single frame of reference, this result could not be predicted. One global reference frame should not result in any difference at all (cf., [12]) or participants should perform better when aligned with the orientation of this single global reference frame as predicted by reference axis theory ([28], [33]). No evidence for this could be observed. Participants seem to encode multiple local reference frames for each vista space in the orientation they experienced this vista space (which coincided with its geometry).

3.3 Egocentric and Allocentric Reference Frames

The reference frames in the network of reference frames theory correspond to vista spaces and they are connected via perspective shifts. Are these relations egocentric or allocentric? Egocentric and allocentric reference frames have been discussed intensively over the last few years (e.g., [28], [46]). In an egocentric reference frame locations and orientations within an environment are represented relative to the location and orientation of a navigator's body in space [15]. This is best described by a polar coordinate system. An allocentric reference frame is specified by a space external to a navigator. Here object-to-object relations are represented in contrast to the object-to-body relations in the egocentric reference frame. An allocentric reference frame is best described by a Cartesian coordinate system.

In principle, the network of reference frames theory is compatible with egocentric as well as allocentric reference frames. With egocentric reference frames, elements within a vista space are encoded relative to the origin of the egocentric reference frame by vectors (and additional rotations if the relative bearing matters). Perspective shifts are just egocentric vectors which point to another reference frame instead of an object within the vista space. Despite in principle being compatible with egocentric reference frames, the network of reference frames theory is better classified as allocentric. This decision is based on five arguments: (1) The origin which is quite prominent in polar coordinate systems does not play a role in the network of reference frames theory. No performance differences are predicted whether a navigator is located at the origin or at another location within a vista space reference frame. A polar coordinate system would suggest that this makes a difference. (2) Contrary to the origin, the orientation of a reference frame does make a difference according to the network of reference frames. When aligned with this orientation, participants should perform better and do so (see 3.2). Such an orientation, however, is more prominent in Cartesian coordinate systems, than it is in polar coordinate systems. (3) The orientation of a reference frame originates either from the initial experience with a vista space or from the vista space's main geometric orientation, (e.g., the orientation of the longer walls of a room). In principle, the main geometric orientation might never have been experienced directly, (i.e., a navigator was never aligned with the surrounding walls). Still, the geometry might determine the orientation of the reference frame (cf., [33]). Although this is a highly artificial situation, such a reference frame has to be allocentric. (4) Within a vista space reference frame, the geometry of the boundaries of this vista space is encoded. It has been shown that the room geometry is encoded as a whole (i.e., allocentrically not by egocentric vectors; e.g., [46]). So at least some of the relations within a vista space are allocentric anyway. (5) Although perspective shifts can be understood as egocentric vectors (plus rotations), they are intuitively better described as relations between locations in an environmental space, (i.e., allocentric relations), rather then relations between egocentric experiences. In summary, the arguments suggest that the network of reference frames theory is better understood as allocentric than as egocentric.

3.4 The Relation between Vista Space Reference Frames: Network vs. Hierarchy

Hierarchic theories of spatial memory have been very prominent (e.g., [4], [11], [40], [50]). In such views, smaller scale spaces are stored at progressively lower levels of the hierarchy. Contrary to these approaches, the network of reference frames theory does not assume environmental spaces are organized hierarchically, but assumes environmental spaces are organized in a network. There is no higher hierarchical layer assumed above a vista space. All vista spaces are equally important in that sense. This does not exclude vista spaces themselves from being organized hierarchically.

Hierarchical graph models or hierarchical cognitive maps still face most of the problems discussed in 3.1. However, one argument for hierarchical structuring is based on clustering effects. In clustering effects, judgments within a spatial region are different from judgments between or without spatial regions. For instance, within a region distances are estimated faster and judged being shorter or locations are remembered lying more to the center of such a region than they were seen before. Many of these clustering effects have been examined for regions within a vista space or a whole country usually learned via maps (e.g., [40]). They are, therefore, not relevant here. However, clustering effects are also found in directly experienced environmental spaces. Experiments show that distance judgments [11] and route decisions between equal length alternatives [49] are influenced by regions within the environmental space. These effects cannot be explained by the network of reference theory alone. A second categorical memory has to be assumed which represents a specific region (cf., [13]). Judgments must be based at least partially on these categories and not on the network of reference frames only. These categories might consist of verbal labels such as "downtown" [22]. As a prediction, no clustering effects for directly learned environmental spaces should be observed when such a category system is inhibited, (e.g., by verbal shadowing).

3.5 Asymmetry in Spatial Memory

The perspective shifts assumed by the network of reference frames theory are not symmetric. They always point from one vista space to another and are not inverted easily. Tasks accessing a perspective shift in its encoded direction should be easier and more precise than tasks that require accessing the perspective shift in the opposite direction - at least as long as there is no additional perspective shift encoded in the opposite direction. This asymmetry can explain the route direction effect in spatial priming and different route choices for wayfinding there and back.

After learning a route presented on a computer screen in only one direction, recognizing pictures of landmarks is faster when primed with a picture of an object encountered before the landmark than when primed with an object encountered after the landmark (e.g., [14]). According to the network of reference frames theory the directionality of perspective shifts speeds up activation spread in the direction the route was learned. Therefore, priming is faster in the direction a route was learned.

Asymmetries are also found in path choices. In a familiar environment, navigators often choose different routes on the way out and back (e.g., [39]). According to the network of reference frames theory, different perspective shifts usually connect vista spaces on a route out and back. Due to different connections, different routes can be selected when planning a route out compared to planning the route back.

The network of reference frames theory explains asymmetries on the level of route knowledge. However, it also predicts an asymmetry in survey knowledge. Learning a route mainly in one direction should result in an improved survey performance, (i.e., faster and more precise pointing), in this direction compared to the opposite direction. This yet has to be examined.

4 Conclusions

The network of reference frames theory is a synthesis from graph representations and cognitive maps. It resolves problems that exist in explaining the orientation behavior of human and non-human animals based on either graphs, maps or both of them in parallel. In addition, the theory explains the unique role of vista spaces as well as asymmetries in spatial memory. New predictions from the theory concern, first, the role of orientation within environmental spaces, which has been tested recently, second, the lack of clustering effects in environmental spaces based on the assumed memory alone, and third, an asymmetry in survey knowledge tasks. Further experiments have to show whether the network of reference frames theory will prove of value in these and other cases.

Acknowledements. This research was supported by the EU grant "Wayfinding" (6th FP - NEST). I would like to thank Heinrich Bülthoff for supporting this work, Bernhard Riecke, Christoph Hölscher, Hanspeter Mallot, Jörg Schulte-Pelkum and Jack Loomis for discussing the ideas proposed here, Jörg Schulte-Pelkum for help with writing and Brian Oliver for proof-reading.

References

1. Bennett, A.T.D.: Do animals have cognitive maps? Journal of Experimental Biology 199, 219–224 (1996)
2. Byrne, P., Becker, S., Burgess, N.: Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychological Review 114, 340–375 (2007)
3. Cheng, K., Newcombe, N.S.: Is there a geometric module for spatial orientation? Squaring theory and evidence. Psychonomic Bulletin & Review 12, 1–23 (2005)
4. Chown, E., Kaplan, S., Kortenkamp, D.: Prototypes location, and associative networks (PLAN): Towards a unified theory of cognitive mapping. Cognitive Science 19, 1–51 (1995)
5. Ekstrom, A., Kahana, M., Caplan, J., Fields, T., Isham, E., Newman, E., Fried, I.: Cellular networks underlying human spatial navigation. Nature 425, 184–187 (2003)
6. Fujita, N., Klatzky, R.L., Loomis, J.M., Golledge, R.G.: The encoding-error model of pathway completion without vision. Geographical Analysis 25, 295–314 (1993)
7. Gallistel, C.R.: The organization of learning. MIT Press, Cambridge (1990)
8. Hamilton, D.A., Driscoll, I., Sutherland, R.J.: Human place learning in a virtual Morris water task: some important constraints on the flexibility of place navigation. Behavioural Brain Research 129, 159–170 (2002)
9. Hegarty, M., Waller, D.: Individual differences in spatial abilities. In: Shah, P., Miyake, A. (eds.) The Cambridge Handbook of Visuospatial Thinking, pp. 121–169. Cambridge University Press, Cambridge (2005)
10. Hein, A., Held, R.: A neural model for labile sensorimotor coordination. In: Bernard, E.E., Kare, M.R. (eds.) Biological prototypes and synthetic systems, vol. 1, pp. 71–74. Plenum, New York (1962)
11. Hirtle, S.C., Jonides, J.: Evidence of hierarchies in cognitive maps. Memory & Cognition 13, 208–217 (1985)
12. Holmes, M.C., Sholl, M.J.: Allocentric coding of object-to-object relations in overlearned and novel environments. Journal of Experimental Psychology: Learning, Memory and Cognition 31, 1069–1078 (2005)
13. Huttenlocher, J., Hedges, L.V., Duncan, S.: Categories and particulars: prototype effects in estimating spatial location. Psychological Review 98, 352–376 (1991)
14. Janzen, G.: Memory for object location and route direction in virtual large-scale space. The Quarterly Journal of Experimental Psychology 59, 493–508 (2006)

15. Klatzky, R.L.: Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In: Freska, C., Habel, C., Wender, K.F. (eds.) Spatial cognition - An interdisciplinary approach to representation and processing of spatial knowledge, pp. 1–17. Springer, Berlin (1998)
16. Kuipers, B.: The spatial semantic hierarchy. Artificial Intelligence 119, 191–233 (2000)
17. Loomis, J.M., Klatzky, R.L., Golledge, R.G., Philbeck, J.W.: Human navigation by path integration. In: Golledge, R.G. (ed.) Wayfinding behavior, pp. 125–151. John Hopkins Press, Baltimore (1999)
18. MacFarlane, D.A.: The role of kinesthesis in maze learning. University of California Publications in Psychology 4 277-305 (1930); (cited from Spada, H. (ed.) Lehrbuch allgemeine Psychologie. Huber, Bern (1992)
19. McNaughton, B.L., Leonard, B., Chen, L.: Cortical-hippocampal interactions and cognitive mapping: A hypothesis based on reintegration of parietal and inferotemporal pathways for visual processing. Psychbiology 17, 230–235 (1989)
20. Mallot, H.: Spatial cognition: Behavioral competences, neural mechanisms, and evolutionary scaling. Kognitionswissenschaft 8, 40–48 (1999)
21. Meilinger, T., Hölscher, C., Büchner, S.J., Brösamle, M.: How Much Information Do You Need? Schematic Maps in Wayfinding and Self Localisation. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, D.R. (eds.) Spatial Cognition V, pp. 381–400. Springer, Berlin (2007)
22. Meilinger, T., Knauff, M., Bülthoff, H.H.: Working memory in wayfinding - a dual task experiment in a virtual city. Cognitive Science 32, 755–770 (2008)
23. Meilinger, T., Riecke, B.E., Bülthoff, H.H.: Orientation Specificity in Long-Term-Memory for Environmental Spaces (submitted)
24. Moeser, S.D.: Cognitive mapping in a complex building. Environment and Behavior 20, 21–49 (1988)
25. Montello, D.R.: Spatial orientation and the angularity of urban routes: A field study. Environment and Behavior 23, 47–69 (1991)
26. Montello, D.R.: Scale and multiple psychologies of space. In: Frank, A.U., Campari, I. (eds.) Spatial information theory: A theoretical basis for GIS, pp. 312–321. Springer, Berlin (1993)
27. Montello, D.R., Pick, H.L.: Integrating knowledge of vertically aligned large-scale spaces. Environment and Behavior 25, 457–484 (1993)
28. Mou, W., Xiao, C., McNamara, T.P.: Reference directions and reference objects in spatial memory of a briefly viewed layout. Cognition 108, 136–154 (2008)
29. O'Keefe, J., Burgess, N.: Geometric determinants of the place fields of hippocampal neurons. Nature 381, 425–428 (1996)
30. O'Keefe, J., Nadel, L.: The hippocampus as a cognitive map. Clarendon Press, Oxford (1978)
31. Poucet, B.: Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. Psychological Review 100, 163–182 (1993)
32. Restat, J., Steck, S.D., Mochnatzki, H.F., Mallot, H.A.: Geographical slant facilitates navigation and orientation in virtual environments. Perception 33, 667–687 (2004)
33. Rump, B., McNamara, T.P.: Updating Models of Spatial Memory. In: Barkowsky, T., Knauff, M., Ligozat, G., Montello, D.R. (eds.) Spatial Cognition V, pp. 249–269. Springer, Berlin (2007)
34. Schnapp, B., Warren, W.: Wormholes in virtual reality: What spatial knowledge is learned for navigation? In: Proceedings of the 7th Annual Meeting of the Vision Science Society 2007, Sarasota, Florida, USA (2007)

35. Sholl, J.M., Kenny, R.J., DellaPorta, K.A.: Allocentric-heading recall and its relation to self-reported sense-of-direction. Journal of Experimental Psychology: Learning, Memory, and Cognition 32, 516–533 (2006)
36. Siegel, A.W., White, S.H.: The development of spatial representations of large-scale environments. In: Reese, H. (ed.) Advances in Child Development and Behavior, vol. 10, pp. 10–55. Academic Press, New York (1975)
37. Skaggs, W.E., McNaughton, B.L.: Spatial Firing Properties of Hippocampal CA1 Populations in an Environment Containing Two Visually Identical Regions. Journal of Neuroscience 18, 8455–8466 (1998)
38. Stankiewicz, B.J., Legge, G.E., Mansfield, J.S., Schlicht, E.J.: Lost in Virtual Space: Studies in Human and Ideal Spatial Navigation. Journal of Experimental Psychology: Human Perception and Performance 37, 688–704 (2006)
39. Stern, E., Leiser, D.: Levels of spatial knowledge and urban travel modeling. Geographical Analysis 20, 140–155 (1988)
40. Stevens, A., Coupe, P.: Distortions in judged spatial relations. Cognitive Psychology 10, 422–437 (1978)
41. Thorndyke, P.W., Hayes-Roth, B.: Differences in spatial knowledge acquired from maps and navigation. Cognitive Psychology 14, 560–589 (1982)
42. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. MIT Press, Cambridge (2005)
43. Tolman, E.C., Ritchie, B.F., Khalish, D.: Studies in spatial learning. I. Orientation and the short-cut. Journal of Experimental Psychology 36, 13–24 (1946)
44. Touretzky, D.S., Redish, A.D.: Theory of rodent navigation based on interacting representations of space. Hippocampus 6, 247–270 (1996)
45. Trullier, O., Wiener, S.I., Berthoz, A., Meyer, J.-A.: Biologically based artificial navigation systems: Review and prospects. Progress in Neurobiology 51, 483–544 (1997)
46. Wang, F.R., Spelke, E.S.: Human spatial representation: insights form animals. Trends in Cognitive Sciences 6, 376–382 (2002)
47. Wang, R.F., Brockmole, J.R.: Simultaneous spatial updating in nested environments. Psychonomic Bulletin & Review 10, 981–986 (2003)
48. Werner, S., Krieg-Brückner, B., Herrmann, T.: Modelling Navigational Knowledge by Route Graphs. In: Habel, C., Brauer, W., Freksa, C., Wender, K.F. (eds.) Spatial Cognition 2000. LNCS (LNAI), vol. 1849, pp. 295–316. Springer, Heidelberg (2000)
49. Wiener, J., Mallot, H.: Fine-to-coarse route planning and navigation in regionalized environments. Spatial Cognition and Computation 3, 331–358 (2003)
50. Yeap, W.K.: Toward a computational theory of cognitive maps. Artificial Intelligence 34, 297–360 (1988)