Learning Minimal Representations for Visual Navigation

William M. Warren, Michael Tarr, Leslie Pack Kaelbling, Andrew Duchon

May 6, 1997

1 Introduction

Emphasize learning throughout.

Navigation is a fundamental ability of animal life. For simpler organisms, this means constant exploration and reaction to the current situation. However, for more advanced organisms, navigation also implies the ability to return to a place, e.g. for shelter, food, or mating. In addition to animals, we might also consider mobile robots for which we, as their creators, desire the ability to do errands and go from one place to another. Thus, we can more generally speak of agents, animate, mobile entities who must navigate through a complex world. The aim of this proposal is to investigate the minimal representations required by agents for visual navigation.

Although all animals that have evolved in the light have developed means of using visual information for navigation in the world, one possiblity of course is that vision is not required at all. Indeed, blind rats () and congenitally blind humans () can learn about their environments. Vestibular and kinesthetic information (together called proprioceptive information here) can provide sufficient information for dead-reckoning–the ability to return to a spot from whence one started (). Proprioceptive information can also be used to learn route sequences (), e.g., go forward 3 steps, turn right, go forward 4 steps, etc.

Visual information, however, provides a clear advantage. Many spatial relations can be perceived simultaneously or in very quick succession. Dead-reckoning, on the other hand, requires temporal integration. Vision provides distal information, and the transformation of distal relations, whereas proprioception provides information about only proximal transformations. Indeed, some studies () show that in humans, early experience with visual information is required for people to gain from dead-reckoning an accurate sense of spatial (distal) relations.

The role of vision in on-line control of action has been explored in a number of agents, including bees (), humans (), and robots (). To explain and produce this behavior, artificial intelligence, cognitive science, and neuroscience in the last 10 years have moved away from the core assumption that symbolic representations and model-based planning provide the necessary basis for intelligent behavior. Work in behavior-based roboitcs has converged with research on biological visual control to demonstrate that sophisticated adaptive behavior can be exhbitied by an agent that is equipped merely with simple task-specific perception-action mappings. Longer-range "cognitive" behavior that depends on information beyond the current sensory horizon seems, once again, to implicate internal representations and planning. In the case of navigation, there are some fundamental questions that need to be addressed concerning the nature of these internal represenations. That is, what is the role of "cognition" in navigation?

There has been a great deal of experimental work asking questions about human and animal navigational abilities (see a brief survey in appendix **??**). Although there are many interesting results, the experiments have not been carried out in a systematic manner, making it difficult to compare the results and draw broad conclusions. We propose to address this problem by doing a set of systematic experiments with humans in a virtual-reality environment. Our experimental set-up will allow a broad range of questions to be asked and answered, yielding data that will provide a sound basis for understanding the algorithms and representations underlying human navigation.

Many researchers have developed theories of spatial representations and navigational methods (refer to something?); these theories range from completely abstract to nearly implementable. We propose to test concrete hypotheses about algorithms and representations by implementing them in a mobile robot and testing the robot on navigational tasks similar to those carried out by the human subjects in the virtual-reality experiments.

The strengths of this project will be in the cross-fertilization between human and and robotic experiments and between cognitive and computer scientists. Computational considerations will give rise to hypotheses about relative difficulty of certain spatial learning problems that can be tested with human experiments. Human experiments will give rise to concrete computational theories that can be implemented and tested on robots.

Unprecedented. You have to give us this money.

2 Research Plan

Based on previous experimental and theoretical work, we can identify several broad classes of navigation strategies that implicate different types of knowledge about the environment:

Path integration (dead reckoning) Based on proprioceptive input about metric changes in position and orientation (e.g. odometry, vestibular and somatosensory input, optic flow), two memory registers, and an integration process.

Image-based navigation Based on visual input about vistas or landmarks, and associative memory for sequences of images. Must see next image from current location; shortcuts can be based on goal vista or reference object; self-intersecting routes are problematic.

Route-based navigation Based on input about current location and associative memory for location-action pairs. Current location may be defined by odometry (take 10 steps and turn right) or visual input about vistas or landmarks (turn right at the tree) or junctions (turn right at the third intersection). Shortcuts can be based on views of intermediate locations; self-intersecting routes are problematic.

Survey-based navigation Based on input about current location and a 2D representation that may have different degrees of structure, e.g.

- Topological: a graph of nodes and links with only ordinal information;
- Weak metric: a graph representation with coarse distance and/or angle information;
- Euclidean: a "cognitive map" with metric distance and angle information. May contain vista, landmark, or junction information that identifies nodes and links.

Shortcuts are based on 2D knowledge which may be qualitative or metric; self-intersecting routes are not a problem.

The general aim of the proposed research is to examine how navigation strategies and the underlying knowledge representation interact with the navigational task and the structure of the environment during learning. Parallel experiments will be conducted on humans in virtual environments and robots in a laboratory environment. We expect that the interplay of these two lines of research will allow us to

- empirically derive models of human navigation that suggest effective strategies for implementation in robots,
- synthetically test possible models of human performance using a robot platform, and
- develop robust map-learning and navigation algorithms for mobile robotics.

To the extent that common task constraints apply to both human and robot navigation, structurally similar solutions may be exploited, although their computational implementation may be quite different.

2.1 Questions

The proposed experiments are motivated by basic questions about how these navigation strategies and the minimal knowledge necessary to support them are shaped by the fact that the agent is embedded in an environment during learning. These questions can be summarized as follows.

1. Representation: What strategies are commonly exhibited in human navigation? What is the structure of the underlying knowledge representation? How does it change with experience in an environment (e.g. from route-based to survey-based)?

2. Exploration: What exploration strategies are used to learn a new environment? How does the navigational task during exploration (e.g. random search, directed search, repeated path, multiple paths) affect the structure of the representation?

3. Environmental structure: How do the structure, regularity, and homogeneity of the environment during learning affect the resulting representation and navigational strategies? For example: (a) Does the preferred strategy (e.g. route-based vs. survey-based) depend on the homogeneity of landmarks, junctions, or path geometry during learning? (b) Does the representation depend on the degree of metric structure in the environment during learning? (c) Do spatial or temporal regularities induce rules that are used to guide subsequent navigation?

4. Metric information: Does the availability of metric information (e.g. odometry, proprioception) faciliate learning or affect the structure of the representation? Does active control as opposed to passive observation facilitate learning, and if so, why (e.g. provides location-action associations, an efferent signal, or guidance of attention)? How are distance and angle represented, and what visual information influences them (e.g. optic flow rate, density of landmarks)?

5. Image-based information: What properties of vistas and landmarks are used in navigation, and is this knowledge similar to that used in object recognition? Specifically, is the representation view-dependent, and does this vary with the distinctiveness of landmarks? At what level of specificity are landmarks encoded, e.g. individual exemplars (this house), subordinate-level (a white ranch house) or basic-level (a house) categories? Are landmarks remembered independently or embedded in the vista context?

2.2 Human Experimental Environment

The proposed human experiments will employ virtual reality technology to establish a general framework within which these questions can be approached in a unified manner. Computer-generated virtual environments will be presented in a head-mounted display, and updated in real time as the subject physically walks around a large volume (a gymnasium) via a motion sensing system. The subject's path in the virtual environment will be recorded by the same system. The advantages of such a VR set-up are three-fold: First, it allows us to address a range of questions about learning and navigation, which have previously been approached piece-meal, within a common methodology. Second, it allows us to flexibly manipulate environmental structures and image information to test specific hypotheses, in ways that would be impossible in a real environment. But most important, it allows us to assess performance directly during ongoing navigation, rather than attempting to infer navigational knowledge indirectly through other tasks. Nearly all previous research has used sketch-maps or models, pointing toward a remembered location, or distance estimation as dependent measures, all of which involve task demands and skills that are quite different from actual navigation. In some experiments we will also collect these measures to assess their validity. Of course, using a VR methodology requires a validation study to insure that performance is comparable to that during navigation in a real environment. For this purpose, we will set up a test environment with screens and landmark objects in the gymnasium and program an identical virtual environment, and compare several navigational tasks in the two domains.

A database of several basic environments will be created, including

Maze World: composed of enclosed hallways that severely restrict the subject's view,

Web World: a landscape composed of places linked by traversible paths with larger vistas, and

Flatland: an open landscape of places and barriers without defined paths.

The properties of the environment will be varied across conditions or during learning in order to test hypotheses about the situation-dependence of the resulting representation, including such variables as

- the number and homogeneity of landmarks,
- the number and homogeneity of junction angles or path lengths,
- trial-to-trial variation in junction angle or path length during learning,
- number of landmarks and junctions within a vista,
- presence of a visible reference object, and
- presence of a "home" location.

An experiment will typically involve a small number of subjects in multiple sessions. The general design of an experiment will include:

1. Exploration phase, in which subjects learn a novel environment $E1$ in multiple sessions by performing various orienting tasks, which manipulate the routes taken, the views seen, and the environmental properties attended to. Orienting tasks, with increasing degrees of constraint, include
   - free exploration,
   - a "scavanger hunt" for a set of goals defined by particular features or views,

- finding a route to one or more goals from the home position,
- finding home from one or more starting positions, and
- learning a specified route linking a series of goals.

The duration of the exploration phase will also be varied to manipulate the subject's level of expertise.

2. Test phase, in which subjects are tested on new navigation tasks in the same $E1$. Navigation tasks include

   - finding a new or previous goal from home,
   - finding home from a new or previous starting position, and
   - finding a new or previous goal from a new starting position (triangle test).

3. Transfer phase, in which subjects perform such navigation tasks in an altered or distorted environment $E2$, designed to assess the knowledge acquired during the exploration phase. Transformations from $E1$ to $E2$ include

   - an affine stretch of the environment that preserves topological structure but not metric distance,
   - an affine shear of the environment that preserves topological structure but changes junction angles and diagonal distances,
   - a stretch or shear of landmarks alone that reassignes them to new locations,
   - local manipulation of distances or junction angles, and
   - change in landmark features, orientations, category level, or number.

4. Non-navigational follow-up tasks wil be conducted in some experiments, e.g. drawing sketch-maps or landmark recognition tests. Data on the paths taken during all three phases of the experiment will be stored on disk and analyzed computationally. Measures will include

   - number of nodes, links, and Euclidean distance traversed from start to goal,
   - stochastic analysis of choices made at junctions, and
   - contingency analysis of choices in the transfer phase to test specific hypotheses (e.g. graph location vs. landmark information).

Within this framework, we can perform almost any manipulation in order to determine its effect on navigational strategy and infer properties of the underlying representation, based on direct measures of navigation behavior.

2.3 Virtual Reality Apparatus

Head-mounted display, etc., etc.

2.4 Robot Apparatus

We will use a human-scale mobile robot equipped with infrared proximity sensors, ultrasonic range sensors, and binocular color cameras mounted on a pan-tilt head. The robot has two fast on-board computers, allowing it to drive through complex environments without requiring cables or a radio connection to host computers.

In previous work, Kaelbling's group has developed a suite of primitive behaviors for navigating in hallways using the ultrasonic and infrared sensors. Duchon has implemented control algorithms based on human and animal performance [?] using optical flow information for obstacle avoidance. This work will be extended to develop a set of primitive navigational behaviors based primarily on vision, but using the ultrasonics and infrared for emergency obstacle avoidance to compensate for limited visual field. Possible primitives include:

- following a wall on the left or right
- driving down the middle of a hallway
- going around an obstacle
- driving to a visible object
- driving in a particular direction (based on dead reckoning)

Robotic vision will also be used for landmark recognition and detection. Tarr's work on object recognition and the integration of multiple views [?] will be applied to developing some visual primitives for use in constructing larger learning and navigation experiments. Examples include:

- determining the possible paths to take (when in an environment with obvious choice-points at corridor or trail junctions)
- visual recognition primitives: stereo? color histograms? Help, Mike!

Robotic experiments will take place in the same large open space as the human experiments. The robot will drive through an environment constructed with partitions and a variety of configurations of visual landmarks.

3 Proposed Experiments

3.1 Real-world to Head-Mounted Display to Big-Screen Validation

See if humans do the same thing in the physical environment as they do when wearing the headmounted display as they do when looking at a screen.

3.2 State and Ambiguity

Mazes are easy when each junction is unambiguously recognizable and get harder as they get more ambiguous (and require history to disambiguate). Even harder are "temporal mazes," where the same location may require different actions depending on how many times it has been visited previously, or other arbitrary properties of previous behavior.

3.3 Metric versus Topological Representations

3.4 Visual Properties of Landmarks

3.5 Exploration

What's the character of free exploration? How do exploration strategies change as the task changes?

3.6 Active versus Passive

How do things change when the agent is active versus passive during the learning phase? How is this affected by the presence of vestibular information?

3.7 Induction of Structural Principles

Can the agent learn regularities in the environment and take advantage of them while navigating new environments (or previously-unseen parts of the same environment)?

4 Proposed Computational and Theoretical Models

5 Educational Impact

We will train graduate students. Give post-docs exposure to exciting interdisciplinary area. Hire lots of UTRAs and REUs.

6 Budget

- 3 post-docs
- 3 grad RAs
- 1 technician (research programmer; robot maintainer)
- hi-res head-mounted VR display
- long-range motion sensing system

- robot upgrade (including compass)
- partitions to make mazes for experiments
- computers
- 1 summer month each for lpk and mjt

Explain that whw has other salary support.

A Background Material

A.1 Animal and Human Navigation

What is the role of cognition in navigation?

That role could be to provide a substrate upon which spatial information could be placed. For example, an agent might go from H to A and from H to B (Figure 1). If they can move directly from A to B (in a sense, perform vector subtraction) we might say that have a "cognitive map" which allows the agent to automatically make a geometric association between two places which have not been sequentially experienced. If one finds such a result, however, one cannot assume such a substrate. For example, if during the test, the agent went from A to H to B, they might simply use dead-reckoning to go from B to A. Or, as we'll see below, when the agent goes from B to A, it might simply be going towards some visual information associated with A that it can see from B. That is, the agent has a mechanism which requires memory (say of an image), but uses perceptual, not cogntive, processes to guide behavior. It can only be successful while acting in the same "vista" (). Numerous other possibilities might lead an agent to apparently metric, accurate behavior. For example, there may be only two paths from B, so on a given trial it has a 50% chance of going to the correct location. There may be a sign pointing in the right direction. It may be able to see H down one path, so try the other in order to reach A. Thus, the ability to take a new path (a shortcut) does not imply that an agent has a Euclidian representation of its environment. To come to such a conclusion first requires the elimination of dead-reckoning and perceptual explanations of this behavior (Duchon, in preparation).

Precisely that conclusion has been reached by most theorists though. Ever since Tolman () first coined the term and posited the existence of a "cognitive map" to explain his results with rats, the idea of a geometric isomorphism has dominated the interpretation of animal and human spatial behavior. Shemyakin () proposed that humans start with "route-maps" but are capable of learning "survey-maps" of their environments. Siegel and White () proposed that children first learn about landmarks, then routes, and finally survey-map representations. Piaget and Inhelder () theorized that spatial knowledge develop from egocentric, topological relations to allocentric, Euclidian ones. O'Keefe and Nadel () proposed that the hippocampus was the seat of the rat's cognitive map, or "locative system" as opposed to a "taxon system" which was only capable of simple S-R associations. Gould's () studies suggested that even honeybees had cognitive maps. Gallistel () gives a more recent ariculation of this theory with regards to rats and bees.

The question becomes, are animals and people really that...

Let us first consider the case of cognitive maps in bees. Dyer () repeated Gould's () experiments on the vector subtraction task and found that bees could not take the short cut when they could not see visual cues associated with the goal location. Other researchers have obtained similar results (). One theory for how the bees do navigate () states that bees store an image of the goal and perhaps a few other images for a route and by moving to match currently visible features with the stored image, they are capable of reaching the goal. This same mechanism can work at various scales () and bees and wasps may even have evolved certain visuo-motor behaviors to ensure that these perceptual mechanisms will work ().

As for rats, it is interesting to note that the original detour () and "sunburst" () experiments that inspired Tolman could not be replicated (, see also Cianci() for various criticisms). Indeed, as the tasks moved from complicated "Hampton Court" mazes to the simpler plus mazes (), the theories became more "cognitive," but with the result that the simple associationists won out (). Maier's three-table task () is probably the closest thing to the vector subtraction task. He found that rats could not take the short cut if the path had not been traversed at least in one direction.

Modern electrophysiological studies of the hippocampal formation have mostly taken place within a single vista. Typical experiments have the rat chasing pellets in a cylinder () or finding food at the ends of a radial maze where the rat must find food at the ends of arms radiating from a central platform (). Both kinds of experiments take place with a single set of "extra-maze cues" surrounding the arena. Thus, it is not clear whether the place fields of hippocampal pyramidal cells can make real spatial inferences across regions the rat has not actually experienced (see Thinus-Blanc, 1995, p. 260). This is also indicated by the range of behaviors which can affect these cells (). When recordings are made in an environment with more than one vista, the results can be ambiguous ().

Modern behavioral studies have mostly used the Morris water maze where the rat must find a platform hidden in a pool of water () and radial mazes (). Even though the hippocampus is necessary to learn these tasks (), it is not clear that the rat is doing anything more sophisticated than the bee in a task similar to the water maze (). Bees have not been tested in a radial maze. Other two vista experiments....(???). Overall then, despite the intense theorizing, there is not overwhelming data on either the capabilities of rats to perform spatial inferences, or on the role the hippocampus might play more than for path integration (). More recent, neural network derived theories will be discussed below.

Human cognitive mapping is a diverse field with studies looking at dead-reckoning in the blind and sighted, the routes and sketchmaps of children, college, students and taxi drivers, and preferences for shopping malls, housing, and parking garages. The list goes on. To limit the discussion, we will present below only evidence that suggests that humans have significant distortions in their mental representations of spatial structure.

As with the early evidence with bees and rats, it is often difficult to tell what the experience of the person has been in his or her environment, what information they have and therefore the inferences they need to perform a certain task. Also, it should be noted that with animals, only behavioral data can be obtained (though the bee's dance and the place cells of the rat's hippocampus give some indication of what they are "thinking"). Most human studies have only looked at more symbolic expressions of knowledge and not actual wayfinding. These generally take the form of sketchmaps, and direction and/or distance judgments from a current or imagined point which are in turn often converted with multidimensional scaling to form a 2-D representations of the subjects' knowledge.

Numerous studies cast doubt on the ability to generalize spatial knowledge across known routes. For example, taxi drivers have clear biases towards learning routes only and are poor at making judgments across them (). Similar difficulties in linking routes have been found by a number of researchers when subjects learn about a new environment (). That is, subjects are relatively accurate at making distance and direction judgments between places on a single route (though going from nearer to farther on the route is easier (?)), but have clear difficulties using the overlap between the routes to make judgments between them. Although subjects can be accurate in making judgments and maps about a building layout when specifically asked to learn such a layout (), users of a building often do not learn more than the routes they need (), much like the taxi drivers. These data suggest that subjects do not automatically learn a Euclidian representation. Moreover, when they do learn Euclidian representations, it may be because they have learned topological constraints. That is, in most cases where people can draw a good sketchmap of a building or city, it may be because they have taken multiple routes through the environment (not to mention the vista relationships that may exist, e.g., a clock tower that can be seen from multiple regions). When these things are controlled for, as Anooshian (1996) suggests, "spatial memories might best be described as a relatively unorganized set of disconnected and independent pieces of information about places within environments." and spatial knowledge is not a "unitary or monolithic entity."

These problems in joining routes may begin at the level of dead-reckoning. When asked to walk two leg of a triangle an point to or return to the starting point, people show systematic biases, either towards the mean (), or, perhaps similarly towards 90° (). This may be quite different than for example the biases that ants show (). The system also seems to break down when subjects cross their own paths (). The bias towards orthogonalizing angles also shows up in sketchmaps and direction judgments in their hometowns (Byrne, Moar and Bower). In many of these studies there are may also be biases from having seen maps of the town however (Kitchin, 1997).

Distance judgments also show biases, this time in terms of hierarchical relationships (). That is, people consider two locations in the same region to be closer than two equidistant locations across regions, whether these regions are defined visually or functionally (). People also tend to segment routes, biasing their distance judgments (). In addition, estimates of walked distance can be biased by the number of turns () and intersections one passes (). All these results emphasize the fact that much more (or less?) than strictly Euclidian information is being represented ().

A.2 Robot Navigation

Needs cleaning up.

There has been a great deal of work on learning maps for mobile robotics and on learning stochastic models of dynamical systems in general. In this section, we focus on map learning for robots.

Different kinds of maps are useful for different purposes and situations. Sometimes it is necessary for a robot to know its location quite accurately in terms of metric coordinates; in such cases, metric maps are clearly the best choice. In many other environments, such as office buildings with corridors and rooms, or networks of roads or freeways, maps that simply specify the topology of important locations and their connections suffice. Such maps are typically less complex and support much more efficient planning than metric maps. Topological maps are supported by lower-level abstractions that allow the robot to move along arcs (perhaps by wall- or road-following) and to recognize at least some properties of the locations.

There are two typical strategies for deriving topological maps: one is to learn the topological map directly; the other is to first learn a geometric map, then to derive a metric map through some process of analysis.

A particularly nice example of the second approach is provided by Thrun and Bücken [**?**], who use occupancy-grid techniques to build the initial map. This strategy makes sense when the primary cues for decomposition and abstraction of the map are geometric. In many cases, however, the nodes of a topological map will be defined in terms of other sensory properties (like labels on a door, for example). Learning a geometric map first also stresses the odometric abilities of a robot; if they are quite weak and the space large, then it will be very difficult to derive a consistent map.

We take the approach of learning the topological map directly, assuming that an abstraction of the robot's perception and action abilities has already been done (we do it by hand, but see work of Pierce and Kuipers [**?**] for an automatic method). This reduces dependence on odometry and allows for states to be defined in much more general terms. Since robots actions are rarely, if ever, completely reliable, map-making has to take uncertainty into account. But a question remains as to what the form of the resulting map should be: it could either be a traditional deterministic map that describes the ground truth of the world, or it could be a combined model of the probabilistic interactions that the agent has with its world (essentially the underlying map with models of uncertain sensing and action). If the stochasticity in the robot's actions and observations is relatively uniform (doors tend to be misperceived in the same way, no matter where they are), then it may be better to learn separate ground truth and error models. In many cases, this is a reasonable assumption. However, we have found that errors are often quite dependent on the details of the robot's location (some locations are much brighter or darker than others; some doors more likely to generate specular reflections in the ultrasonics), so we would like to learn a combined model.

Kuipers and Byun [?] provide a strategy for learning deterministic topological maps. It works well in domains in which most of the noise in the robot's perception and action is abstracted away; it learns from single visits to nodes and traversals of arcs. In addition, it has limited ability to cope with situations in which long strings of actions and observations are necessary to disambiguate the robot's location. Another set of learning algorithms, based on the theory of learning deterministic finite state automata, work in much noisier environments with much less global information. Basye, Dean, and Kaelbling [?] provide algorithms for learning deterministic maps given (fairly strong) assumptions; these algorithms come with probabilistic correctness guarantees for learning in polynomial time with a polynomial amount of data.

The work most closely related to ours is that of Koenig and Simmons [?], who learn POMDP models (which are essentially stochastic topological maps) of a robot hallway environment. They also recognize the impossibility of learning such a model from no initial information; they solve the problem by using a human-provided topological map, together with further constraints on and information about the shared structure of the model. They then use a modified version of the Baum-Welch algorithm to learn the transition and observation probabilities in their model. They also developed an incremental version of Baum-Welch that allows it to be used on-line in certain kinds of environments. Their models learn and use very weak metric information by modeling hallways as chains of 1-meter segments and allowing the learning algorithm to select the most probable length of the chain. This method is effective, but results in very large models.

Also: Ernie Davis; Engelson and McDermott

A.3 Virtual Reality Environments

A.4 Cross-fertilization of Robotics, Psychophysics, and Insect Ethology

A.5 Network models

Over the past three years, Duchon has developed a simulation of the real robot. This simulation has an agent wander through an environment, sampling from the optic flow field. This optic flow information is then used in an identical way as the real robot. A number of methods for navigation have been investigated using this simulation (Duchon, 1996). These included a motor-based bias in the obstacle avoidance control strategy, a perceptuall-guided bias that could learn a route sequence and return home, and a geometrical model using extra-maze cues based on O'Keefe's hippocampal model. Each of these had greater capabilites than the previous one. However, the process of trying to develop a flexible method for navigation revealed the limits of even the geometrical model. Therefore, in the last two years, Duchon has been developing a neural net model based on Anderson's attractor networks (). Although flexible sequence learning has been developed (much like Levy's work ()), actually using this information and being able to learn new information at the same time has proven difficult. Duchon is now at the stage, however, of applying Hasselmo's () more cellular models of hippocampal learning and recall to this problem. Analyses of the relationships between Anderson's networks and Hasselmo's suggests that this could be very fruitful. The overall goal of this work is to produce a recursive attractor network.

Why an attractor network? This is suggested by the results showing hierarchical effects. That is, it can not be strictly view-based, you must be able to chunk views together continuously. However, during transitions, there will be sharp changes, and a new basin of attraction needs to be created. The longer one is in a single attractor though () the deeper it will become and so distance can also be represented. Having extra-maze cues allows these individual attractors (locations) to be represented nearby in the state space–leading to a hierarchy.

Why recursive? The types of route-effects that have been seen in the literature point to this type of solution and can accommodate other models which use a view-action-nextview method (Kuipers, Mallot). Recursive neural networks have recently become popular in dealing with the problems of language, especially the temporal ones (). Connections can be easily drawn between this work and the problems of navigation. An environment has a certain vocabulary (places) and allows only certain kinds of syntax (routes). Simple recursive networks have been shown to be able to learn the latter without explicit representation of rules. This is equivalent to the idea of getting metric behavior out of system that has no "rule" saying a triangle has 180° or three right turns make a left. Just as there as the various relations between words can be represented, so can those between places. But, as in the previous paragraph a place is not a discrete symbol (or view) as in language, and therefore recursive, attractor networks are the most promising.