COGNITIVE SCIENCE **19**, 1–51 (1995)

# Prototypes, Location, and Associative Networks (PLAN): Towards a Unified Theory of Cognitive Mapping

ERIC CHOWN
STEPHEN KAPLAN
DAVID KORTENKAMP

*Department of Electrical Engineering and Computer Science*
*The University of Michigan*

An integrated representation of large-scale space, or cognitive map, called PLAN, is presented that attempts to address a broader spectrum of issues than has been previously attempted in a single model. Rather than examining wayfinding as a process separate from the rest of cognition, one of the fundamental goals of this work is to examine how the wayfinding process is integrated into general cognition. One result of this approach is that the model is "heads-up," or scene-based, because it takes advantage of the properties of the human visual system and, particularly, the visual system's split into two pathways. The emphasis on the human location or "where" system is new to cognitive mapping and is part of an attempt to synthesize prototype theory, associative networks and location together in a connectionist system. Not all of PLAN is new, however. Many of its parts have analogues in one or another preexisting theory. What makes PLAN unique is integrating the various components into a coherent whole, and the capacity of this resulting system to speak to a wide range of constraints. Our approach emphasizes adaptiveness; thus, our focus on such issues as ease of use and efficiency of learning. The result is a model that has a stronger relationship both to the environment, and to the ways that humans interact with it, compared with previous models. The resulting model is examined in some detail and compared to other systems.

## INTRODUCTION

Wayfinding is an important and complex task. The specialized structures that humans use for this task are called cognitive maps (Golledge, 1987). Although cognitive maps are useful for a wide variety of reasons (Kaplan & Kaplan, 1989), their fundamental purpose is wayfinding. A cognitive map serves two functions with regard to wayfinding: representing environments,

---

Correspondence and requests for reprints should be sent to Eric Chown, Department of Computer Science, Dearborn Hall, Oregon State University, Corvallis, OR 97331.

and the corresponding ability to use the representations to move from place to place within the mapped environments. This article presents a representational theory of human cognitive mapping that addresses a wider spectrum of wayfinding issues than most other theories in the literature. Part of what makes this work unique in the cognitive mapping literature is that it draws on solutions from other domains of cognitive science and synthesizes them in a new way. In particular what stands out in PLAN (Prototypes, Location and Associative Networks) is an account of the impact that the capabilities of the human visual system, and in particular its two subsystems (the *what* and *where* systems), have on the kinds of cognitive maps that people ultimately develop.

Human wayfinding can be broken down into four component problems: landmark identification, path selection, direction selection, and creating abstract environmental overviews. Solutions to the first three of these problems are essential to human wayfinding. The fourth problem, achieving environmental abstraction, is a means of substantially increasing efficiency and functionality. Although the four problems are individually separable, a unified solution requires an integrating framework to mesh them together.

> **Landmark identification** is the most basic component of wayfinding. Landmarks are environmental place markers vital in determining orientation and current location. The landmark identification problem in wayfinding is to separate out distinctive objects in the environment, called landmarks, which can later be used in route planning and can be recognized while traversing the chosen route.
>
> **Path selection** involves choosing a route to the goal. In this case a path is not a direction, but is more algorithmic, for example, a series of places that will lead to the goal. In many models of cognitive mappings, paths are conceptualized as sequences of landmarks. To follow a path one goes from landmark to landmark in the sequence; at each landmark in the path it will be necessary to select the next landmark in the sequence.
>
> **Direction selection** involves choosing a direction in which to travel. If the goal is in sight, for example, a reasonable direction to pick would be towards the goal. For goals that are not in sight the direction selection problem is of course more difficult; beyond the fact that the goal cannot be seen, sometimes a journey will require a series of turns and shifts in direction. Thus, direction selection at the starting point is rarely sufficient to guide an entire route.
>
> **Abstract environmental overviews** are a further generalization of the route concept. If one were to travel extensively in a particular environment, it would be useful to have a coherent overview of the entire environment. Rather than dealing with routes individually, such

a structure would allow them to be extracted from a common abstraction. In addition, this overview would make large-scale reasoning about the environmental simpler. However, although these overviews do serve to increase the efficiency of wayfinding, they are not strictly necessary. Thus, we will include a discussion of wayfinding without such capabilities.

Each of these four subproblems has a different character and consequently is likely to require a different solution. The landmark identification problem primarily concerns the object recognition system. The path selection problem is cognitive, often requiring the selection of one path among a number of alternatives. The direction selection problem, while generally visual, is more locational than the landmark identification problem. Finally, the problem of creating an abstract environmental overview requires a hierarchical synthesis of each of the other three solutions. Although we are taking a modular approach, it need not be the case that the subsystems are completely separated; for example, knowing what to expect next on a path may be a useful bit of information when identifying a landmark. Therefore, although we will discuss each problem separately, throughout the article we will be careful to note how the different parts interact in the overall theory.

The solutions offered by PLAN to each of these problems are not necessarily new. Indeed, this model builds upon our own previous work, and that of the connectionist oriented SESAME[1] group. In particular, PLAN represents a synthesis of prior work on associative networks (Kaplan, 1973; Levenick, 1991), on prototypes (Kaplan, Sonntag, & Chown, 1991), and on spatial or locational visual processing (Lesperance, 1990). Thus, what is new is the attempt to construct a coherent synthesis of several previously divergent lines of research. The complementarity of these various components has, in fact, exceeded our expectations. The resulting system has a number of interesting properties, among them economy, simplicity, and error tolerance.

This article begins by laying out the key constraints that have driven the PLAN model. We take for granted that PLAN is a connectionist model; because the portions of our previous work incorporated in PLAN are all connectionist, this constraint is implicit. (Readers interested in the SESAME approach to connectionist modeling can refer to Kaplan, Weaver, and French (1990) and Kaplan et al. (1991).) The remaining constraints can be divided into three catagories: (1) constraints deriving from the fact that cognitive maps were developed in an evolutionary context; (2) physiological

---

[1] SESAME is an acronym for Seminar in Environmentally Sensitive Adaptive MEchanisms. This University of Michigan, multidisciplinary research group has explored and extended Hebbian connectionism over a period of some 20 years.

constraints, particularly pertaining to the brain structure associated with processing spatial information; and (3) what we will call developmental constraints, but which will also subsume general psychological constraints derived from studies of human behavior. In the third section of the article we will present the PLAN model. Although we will compare specific parts of PLAN to pieces of other models along the way, we save extended comparisons until after the entire model has been presented. The fourth section, therefore, will center around a more general discussion of PLAN and how it compares to, and differs from, other theories in the literature.

## CONSTRAINTS

There is an indefinitely large body of material that might be considered at least somewhat pertinent to a theory of human cognitive mapping. It is not our intention to provide a comprehensive analysis of all possible constraints. Rather, our focus is on key themes, on constraints that suggest general principles under which such a model should operate.

### Evolutionary Constraints

Evolutionary constraints concern the incremental and robust development of cognitive maps. It is unlikely that structures as complex as cognitive maps could have developed all at once; rather, they are likely to have developed one piece at a time, each piece contributing in some positive way to the overall functionality of the map. This has a wide range of implications for the types of structures that might develop. Some of these implications will be addressed as they arise throughout the course of the article, but a few are sufficiently central to introduce at the outset: These are simplicity, consistency, and economy.

**Simplicity** of structure is important for a number of reasons. First, simple structures are less likely to break down. The fewer interactions among components, the smaller the chance of error. Second, simple structures are more plausible from an evolutionary standpoint. Complex structures may require parallel development of a number of separate pieces. However, each piece of the structure must be useful in its own right or it would not develop in the first place.

Evolution is essentially a conservative process (Clark, 1989). In terms of the brain, this means that mechanisms and structures that have proven useful in one domain are likely to be used in other domains when possible; change comes out of necessity. Like simplicity, **consistency** will tend to bring reliability, an essential ingredient for evolutionary success. The consistency principle mandates that similar mechanisms and structures will be used in a wide range of domains. The brain is very unlikely to use vastly different representational structures according to the needs of a particular task;

rather, it is likely to have developed general-purpose structures that are widely adaptable.

The economy constraint is closely related to what Clark (1989) called the "**007 principle.**" This principle gets its name from the fictional character James Bond who, as a spy, was only given information on a "need to know" basis. The principle states that creatures will neither store nor process information in costly ways when they can use the structure of the environment and their operations upon it as a convenient stand-in for the information-processing operations concerned. For example, rather than computing and storing the exact coordinates of some object, a more economical solution might be simply to store the fact that it is next to a well-known landmark. Storing and processing information comes with a cost; if creature *A* can get away with less storage and processing than creature *B*, then it has a survival advantage in the long run.

These principles are fairly basic, but it is easy to lose sight of them when examining one particular piece of the larger cognitive mapping puzzle. Perhaps this is because such principles make more sense in the context of the whole cognitive map than when studying the individual pieces. These constraints are useful because they limit the space of possible forms that cognitive maps may take and provide a framework by which unified models can be judged.

## Physiological Constraints

For sighted humans, spatial information characteristically comes from the visual system because it directly affords the richest source of spatial relationships. Directional, or locational, information, however, is distinct from the kinds of information useful for recognizing objects. Indeed, it has been found that there are two separate subsystems within the visual system, one for recognizing objects and one for discerning the spatial locations of objects (Otto, Grandguillaume, Boutkhill, Burnad, & Guigan, 1992; Ungerleider & Mishkin, 1982). These systems are often called the *what* and *where* systems or the *contour* and *location* systems (Kaplan & Kaplan, 1989; Lesperance, 1990; Rueckl, Cave, & Kosslyn, 1988).

These two systems provide humans with two different types of information even though most people are unaware that such distinctions are being made. One system identifies objects and the other determines where things are in space. There is clearly a strong correspondence between these two systems and what we have called the landmark identification problem and the direction selection problem. The landmark identification problem requires object recognition, whereas the direction selection problem requires knowing where those landmarks exist in space relative to each other and to the observer.

It is possible to imagine that cognitive maps could exist without one or the other system. For example, a cognitive map without locational capabilities might simply consist of a network of landmarks. Navigation in such a system would involve moving to a landmark, looking around for the next landmark, and repeating this process until the goal has been reached. Navigating with a cognitive map based upon the locational system, on the other hand, would involve a different mode of operation more akin to what is often described as having a good sense of direction. In such a mode one knows where one is relative to one's goal and individual landmarks along the way are not so important. In fact, we will argue that humans are capable of functioning either way, where one mode of functioning relies heavily upon recognition of nearby objects and the other mode relies more on knowing the relative locations of objects, even when they may be widely separated in space.

A body of research that is closely related to the work on the two visual systems centers around the role of the hippocampus in spatial processing. One physiological theory of cognitive mapping even posits that the function of cognitive maps takes place almost exclusively within the hippocampus (O'Keefe, 1989; O'Keefe & Nadel, 1978). Most of the evidence concerns the relationship between firing rates of neurons within the hippocampus as modulated by different locations and orientations within a location. Such studies suggest that the hippocampus is used to store and process scenes. In particular, O'Keefe speculated that different scenes are stored and can be indexed by the head orientation of the organism as well as its physical location and orientation. He postulated that there are default-stored representations for a particular location and went on to describe how stored scenes might be matched against visual input. More recent evidence presented by Squire (1992) suggests that the hippocampus is more of an interface between stored spatial memory structures and the currently perceived environment. Such an arrangement is not at odds with O'Keefe's findings and suggests that his theory of how the hippocampus resolves perceived versus stored scenes may be sound. However, Squire's evidence also shows that there is a great deal more to spatial functioning than can be accounted for in the hippocampus alone, as evidenced by how organisms with damage to their hippocampus are able to function. Nevertheless, the hippocampal data, especially in conjunction with other data on the physiology of the visual systems, provides important insights into the nature of the kinds of spatial structures that are contained in cognitive maps.

A significant portion of PLAN is based upon this type of locational information, combining what is known about the *where* part of the visual system with the evidence on how the hippocampus functions to provide the basis of a system that extracts basic spatial relationships from scenes and can store and retrieve such scenes based upon using cues such as the orientation of the eyes, head, and body.

## Developmental Theories

If there is one area of research within the field of cognitive mapping that is relatively free of disagreement, it is the work on the developmental sequence of cognitive maps. The aim of this research is essentially to provide functional descriptions of the stages in which children's cognitive maps develop. There is also evidence that cognitive maps in adults show these same stages in new environments (Golledge, 1987). The characterization of the developmental approach that we will present in the following sections is a distillation of the theories of Piaget (Piaget & Inhelder, 1967), Siegel and White (1975) and Shemyakin (1962). Although all three lines of work vary somewhat in terminology and in exact separation of stages, they are in general agreement. One goal of this article is to describe the mechanisms responsible for the developmental sequence, an area that Bates and Elman (1992) considered to be poorly developed in most cognitive theories.

### Landmarks

The first stage of development could be called the object, or landmark, level. As we shall see, this is one area that computational models of cognitive maps have tended to gloss over and it is also the case that developmental theories spend little time on the nature of landmarks. The developmental theorists stress the differentiability of landmarks, but offer little theoretical underpinning as to what constitutes one. For such theorists the important point is that children must learn the objects in the world before they can start putting the objects together in a structured way to form a representation of the world.

As research on computer vision has shown, separating out objects from their environment and recognizing them is far from a simple task. Because this task is so complex and because it is so intimately linked to the context in which we see objects, we will argue that landmarks must be treated as structures far more complex than mere symbols to be plugged into a higher level representation. As the developmental theorists do not delve into such issues, we will turn instead to the psychological literature, which has examined the object recognition problem in considerable detail; our discussion of landmarks will center around the concept of prototypes, a representation that has been proposed for how humans store objects.

### Route Maps

The second developmental stage is usually called the route map stage. At this stage the cognitive map appears to consist of a large collection of routes. Shemyakin (1962) provided one of the best descriptions of this phenomenon when he described how children draw pictures of their neighborhoods. Rather than drawing the neighborhood from a global perspective, children draw it as though they were imagining walking around it, even turning the paper as they turn in their routes. Quite simply, children's cognitive maps closely resemble their experience at this stage. Their experience consists of walking various routes around their neighborhood, and their cognitive maps reflects it. At this stage there is very little conception of the neighborhood as a whole entity. Shemyakin's description is striking because it is so visually and, in particular, locationally oriented: Children turn the paper to correspond with their mental image of what is in front of them.

Piaget, perhaps because he was not looking at navigation, but rather spatial development as a whole, subdivided this stage (Piaget & Inhelder, 1967). The route map stage covers the egocentric and preoperational phases of Piaget's developmental sequence. For Piaget, the child first begins to relate objects to self and later begins to see relationships between objects beginning with topological relationships. This viewpoint is completely consistent with the route map hypothesis. At the route map stage, the information is assimilated directly, reflecting the child's experience rather than an abstract data structure. It is the environment that shapes the representation rather than the representation shaping the perception of the environment. This reliance upon direct experience could be called egocentric. As children move through an environment, the experiential information they receive will be of a topological character. A child need not even move through an environment to receive topological information; it is only necessary to move the eyes. Things that are close together will be experienced close together. The representation reflects the environment by capturing the experience of it.

The differences between Piaget's (Piaget & Inhelder, 1967) characterization of routes and Shemyakin's (1962) harken back to our earlier discussion of two possible types of functioning within a cognitive map: one that relies upon recognizing nearby objects and one that consists of knowing more about the relationships of objects. Piaget's characterization suggests a topologically organized collection of landmarks. A journey using such a structure could be characterized as going from landmark to landmark. Indeed, as we shall see later in this article, such a representation forms the basis for a number of models of cognitive maps. Shemyakin's description is also essentially egocentric, but there are subtle, yet telling, differences. As the children describe their neighborhoods they turn their paper as they would turn if they were walking a route. Such a knowledge structure is not topological, but instead could be characterized as directional. A route in such a structure also consists of going from place to place, but the places no longer necessarily correspond to landmarks, but instead correspond to the actual locations that the children move through in their journeys, particularly those places where a change in direction is necessary.

Tieing the evolutionary, physiological, and developmental arguments together, we are suggesting that there are parallel structures in cognitive maps at this stage. The first structures are like Piaget's (Piaget & Inhelder, 1967) characterization, topological representations of landmarks. Such structures are very simple, and therefore quick, to learn. Such structures also provide a useful means of planning journeys with a relatively minimal amount of information. At the same time a more directional structure is being developed. Because it requires more information, such as information about the relative locations of landmarks, it is slower to develop. However, because it contains this extra information, in the long run it is potentially more useful. The complementarity of such a system is useful from an evolutionary perspective because the organism with both structures would not need to rely on either one, and provides a kind of graceful degradation in case there is damage to one or the other system. Such a system can also account for a great many individual differences in learning and functioning in environments as different people will acquire the two structures at different speeds and might rely upon one or the other to varying degrees. We will discuss these two forms of route maps, which we will call topological and locational, in much greater depth when we present the details of PLAN later.

### Survey Maps

The next stage of cognitive map development, the survey map, is often ignored by computational models of cognitive maps. One tendency of existing models is to conceptualize survey maps as being essentially like cartographic maps. In other words, it is assumed that they capture very precise Euclidian spatial information. Indeed, this is the direction that much of the work in robotics has taken as well. There are, however, little or no data to support this notion in people.

The transition from route maps to survey maps is marked by two important steps. First, an objective frame of reference appears to be developed. To continue with Shemyakin's (1962) example, children at this stage no longer need to turn the paper as they draw the neighborhood. Rather, they sketch whole sections at a time, as if they were able to see them, instead of relying on sequences. It appears, therefore, that they are no longer recreating specific experiences, but are now integrating experiences to form a coherent whole.

Of course if route maps truly are represented as networks, they already reflect some integration across experiences. However, route maps afford only partial spatial information at a local level. In particular, what route maps lack are easily recoverable relationships between distant objects. The second significant step beyond route maps that survey maps take is the ability to determine spatial relationships of objects that are not close in space. The fundamental difference between route maps and survey maps is that in one case the information is local and in the other case it is global. However, as

Shemyakin's (1962) example shows, survey maps may be fundamentally visual and image-like just as the route map structures appeared to be. In returning to the theme that there are two kinds of functionality within cognitive maps, what this suggests is that survey maps are large-scale directional maps that provide a visual overview of a space that is normally too large to be seen all at once (except from an elevated perspective). An essential principle of PLAN will be that many of the structures within a cognitive map, including the survey map representations, will be visual in nature and that processing with such structures will be closely akin to seeing the environment in one's head.

However, although such structures would capture spatial relationships between distant objects, they do not seem to be objective frames of reference, which is supposedly one of the crucial developments in the survey map. Indeed, our claim is that survey maps, rather than being objective, are anchored at a particular perspective. However, as we shall see in the discussion of our own version of survey maps, this is not a problem when the perspective at which the maps are anchored is at the *edge* of the environment. In such cases everything within the environment can be "seen" and the spatial relationships between any arbitrary objects can be determined because everything is in front of the observer. It will also generally be the case that multiple survey maps can be created for a single environment making it possible to switch to the viewpoint most appropriate to a given task.

### Summary

The combination of the evolutionary, physiological, and developmental constraints suggests a basic structure and sequence of development for cognitive maps. This structure starts out with landmarks, distinct objects in the environment. The first stage of mapping consists of learning a topological structure built out of landmarks through experience. Use of such a map would then consist of extracting a sequence of landmarks from one's starting point to one's goal. As experience within the environment grows, an additional type of route map also comes to be. Rather than consisting of a sequence of landmarks, this structure consists of a sequence of places that one has experienced. Corresponding to each of these places, a directional structure is stored that codes relative directions of landmarks. With such a structure in place, the role of the landmark goes from destination to marker. That is to say that a trip is no longer conceptualized as a series of landmarks, but instead it is a series of places, and landmarks are useful in keeping track of where one is. However, even at this stage the information used is local: what can be perceived from where one is standing. Once survey maps are developed, however, then global information is available for wayfinding. With a survey map one can "see" the direction one needs to go to reach one's goal even when it is far away.

Such a system is grounded in the evolutionary principles that we have set forth. Each new stage builds upon the functionality of the previous stage and even the early stages provide the organism with the ability to do useful wayfinding. The two types of routes developed provide the system with useful complementarity and redundancy helping to make it robust. On the physiological side, these two types of routes reflect the two types of processing done in the visual system. The *what* system relies only upon the identification of objects; location is irrelevant. The *where* system, on the other hand, codes the relative location of objects. Because there is more information and more processing involved in determining the relationships of objects, the process of learning such structures is necessarily slower than the simple process of building a topological structure out of landmarks. With enough experience a survey map can be constructed, providing additional functionality and, we will argue, building upon the structures already in place.

In the next section we will describe PLAN in detail. The structure of our description is based upon the developmental sequence already described: landmarks, routes (both topological and locational) and then survey maps. We have already outlined the general constraints that PLAN should meet. It should bring together three distinct pieces: the developmental and psychological literatures, the physiology of the visual system, and a connectionist framework. Once we have described the system in detail we will examine how well it meets these constraints and will compare it to other computational models.

## PLAN

### Landmarks

Landmarks function as a kind of environmental index. Recognizing nearby landmarks is enough to tell one where one is in a familiar environment. Consequently, the fundamental property of landmarks is that they must be uniquely identifiable. Within the cognitive mapping literature, this is almost the only property of landmarks that seems to matter. This is mainly due to the fact that unique identification is a perceptual issue and therefore, apparently, out of the domain of the more high-level issues that cognitive map research typically focuses on. However, this does not mean that there is not an extensive literature on landmarks. On the contrary, it is our position that landmarks form a special case of widely studied issues in cognitive science: object recognition and categorization.

Landmark recognition cannot be treated as a separate issue from object recognition. If anything, landmark recognition is more difficult than object recognition because there is less room for generalization. For example, it is often good enough to recognize that the creature on the hill is a dog, but a landmark must be a particular instance, such as the tree with the big branches close to the ground on one side. Having separate object and landmark systems would require extra bookkeeping and processing. It would violate our economy principle if an object in one context could be a landmark in another. Such a separation also suggests that there is some a priori way of distinguishing between an object and a landmark, which cannot be possible because every object is a potential landmark. Because there is no a priori way to tell the difference between an object and a landmark, there is no reason to have separate representation system. Such an arrangement would be a direct violation of Clarke's (1989) 007 principle. It simplifies matters considerably to have one representation that is useful in both contexts.

The landmark identification problem, therefore, is in reality a special case of the object recognition problem. The object recognition problem is not quite the same as taking a given visual input and deciding what it contains, because it assumes that a pertinent object has already been extracted from the visual field, a major piece of processing in itself. Given that the object has been extracted, the object recognition problem becomes one of categorization; the problem is to determine if a given input corresponds to any known categories. In the case of a cognitive map the categories are the individual landmarks, and the problem when looking at an object is to determine if the object corresponds to any known landmarks.

Treating landmarks as categorical represents something of a departure from standard thinking. On the face of it, a landmark would seem to be a single, unique object, not a category. From the point of view of the sensory input an organism would receive, however, any object can give rise to a multitude of stimulus configurations. Seen from different distances and at different viewing angles, under varying lighting conditions, and even sometimes partially obscured by other objects, a landmark (like any other object) presents a challenge to traditional theories of pattern recognition. The variety of stimulus configurations must somehow be integrated into a whole so they can be recognized as functionally the same thing. Such a process of identifying a variety of different patterns as serving as an equivalence class with respect to a common label or referent is what is usually referred to as categorization.

In summary, the representations that will be used in the identification of landmarks are exactly those used for categorization. The representation of landmarks are constrained in the same ways as the representation of any other categories.

### Prototypes

Prototype theory was developed to address the constraints that a categorization scheme must meet (Posner & Keele, 1968; Rosch, 1978). Prototypes are usually thought of as being the typical example of a category; the prototypic dog, for example, would have four legs, a tail, and so on. Prototypes are generalizations derived from a range of experience. In such a generalization the features that occur most often come to represent the prototype, whereas features that occur less often are weaker, giving the prototype a statistical nature, reflecting experience.

Just because prototypes are generalizations does not mean that individual objects are not represented as prototypes. Indeed, we may experience an individual dog in a wide variety of contexts and see it under varying conditions. In the case of a landmark we can experience it from different angles, distances, orientations, lighting conditions, and the like. An individual representation will be the prototype of a large number of experiences, those features which best represent the landmark forming the core of its representation. So, for example, a building might have a black door on one side, but if that door is not typically seen when looking at the building, then it will only weakly be a part of the prototype. On the other hand, if the door is on the side of the building that is most frequently seen it will be a very important part of the prototype and the building will probably typically be thought of from that orientation.

Prototypes are organized in hierarchies where the levels range from the specific to the abstract. At the bottom of a hierarchy will be individual examples with each successively higher level bringing a corresponding increase in the generality of what is being represented. For example, individual trees will be at the bottom of a hierarchy, but one level up might have different types of trees such as oak or maple, another level could contain classes such as conifer and deciduous, and another level could contain the general category of trees.

Prototype theory has been attacked, in part, due to some confusion resulting from multiple uses of the term (Lakoff, 1987), and to the fact that it is statistically based. Keil and Batterman (1984), however, have shown that young children develop statistical categories before shifting to more theoretical ones. This suggests that prototypes may serve as the basic category structure with theoretical categories emerging from contextual issues. With regard to landmarks, nothing beyond a statistical representation is likely to be necessary because there is no generalization of landmarks. All of the variation is within a landmark, not across landmarks. Whereas "unusual buildings" may serve as a general class of landmarks, the modifier "unusual" denotes that this is a category in which what the individual members have in common is that they are different. After all, the characteristic feature of landmarks is that they must be distinctive. For this reason virtually every model of human cognitive mapping determines what objects in the environment are landmarks by perceptual saliency, a statistical construct. For a discussion of a representational theory of prototypes on which the prototypes used in this article are based, see Kaplan et al. (1991).

### Prototypes and Association

One critical property of prototypes relates to what Bruner (1957) called "going beyond the information given." This refers to the fact that we do not necessarily have to see an entire object to activate its entire representation. For example, seeing a large gray thing with a long trunk is probably enough to make one think of an elephant. This is an important property of object recognition: seeing pieces of an object can be sufficient to activate the correct representation. Furthermore, it means that activating a representation automatically brings with it what Gibson (1979) called *affordances.* In his terms, "The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes,* either for good or ill" (p. 127). These adaptive associations are intrinsic to the network structure developed here. It should be noted that each representation of this system is symbol-like in many of its functions, it also has the capacity for activity and the capacity to activate other representations. For this reason the connectionist theory of prototypes that we will assume here has been called "active symbol theory" (Kaplan et al., 1990).

Of course, not every landmark can be easily distinguished. For example, a particular tree might serve as a landmark. It is essential that seeing the tree activates the representation at the right level of the hierarchy. There must be something to differentiate individual trees if they are to be landmarks. Otherwise, a more general representation may be activated, one which could correspond to many different trees and therefore is not unique to any spatial location. Appleyard (1969), in findings later replicated and extended by Evans, Smith, & Pezdek (1982), found that visual distinctiveness, whether it was through size, shape, or a variety of other influences, is one of the critical factors that determines whether or not a building will be recalled by people familiar with an area.

The Evans et al. (1982) study went beyond Appleyard's (1969) original data and looked at a variety of other factors that influenced recall. One of the additions that added significantly to the predictive power of Appleyard's model was street context. This measured the uniqueness of a building's architectural style on a particular street. For example, a modern building among Victorian houses would be rated high in street context and would be more likely to be recalled. This suggests that the overall structure of the environment is critical in determining what will make a good landmark within that environment. It may not be sufficient just to see a landmark; it may also be necessary for that landmark to be in the proper setting.

### Prototypes and Cognitive Maps

The internal structure and dynamics of the representations of prototypes will have a significant impact upon the higher level structures in cognitive maps. A reasonable question to ask would be: "What is the minimum amount of theory about the internal representations of objects that a model of cognitive mapping must account for?" We propose the following list:

1. Landmarks must be recognizable. This in turn requires that (a) a landmark must be recognizable from a variety of views and orientations, and (b) that in many cases only a partial view of the landmark should be sufficient to activate the entire representation.
2. The number of landmarks which can be active, or processed, at one time is limited to 5 ± 2, the number of objects that a person can think of at one time (Mandler, 1975).[2]
3. Landmarks are intimately linked to context. A good landmark in one environment may be a poor one in another environment. In addition, in a familiar environment the activation of a landmark might not even require seeing it. Conversely, seeing a familiar landmark is often enough to call to mind its setting.

The landmarks in the majority of computational cognitive map theories are based almost exclusively upon the first point. In robot implementations, landmarks are often simulated using simple beacons. Similarly, in computer simulations of cognitive maps, landmarks must generally be assumed, as they are in the Traveller (Leiser & Zilbershatz, 1989) and Tour (Kuipers, 1978), because object recognition is simulated. Even the NX robot, which builds upon Tour (Kuipers & Byun, 1991), constructs its landmarks strictly on the basis of perceptual distinctiveness. Qualnav (Kuipers & Levitt, 1988) uses a similar rule, calling its landmarks "distinctive visual events," and Mataric's system (1990) also relies completely upon sensory characteristics. To some degree, perceptual distinctiveness can include context, for example, "the house next to gas station," but this is quite different than high-level knowledge such as "the fifth traffic light." Furthermore, Kaplan (1976) differentiated *perceptual* distinctiveness from *functional* distinctiveness. Functionally distinctive landmarks serve some useful purpose for an individual and are learned primarily through frequency or repetition. Some systems do at least claim to take into account such issues, such as the Traveller, which acknowledges that a landmark can be an object that is important "in the cognitive scheme of the user," or in Navigator where landmarks are defined both by the perceptual and subjective importance criterion (Gopal, Klatzky, & Smith, 1989; Gopal & Smith, 1990). However,

---

[2] In 1956 there appeared an article by George Miller entitled "The magical number seven, plus or minus two: Some limits in our capacity for processing information." Due to this insightful and often cited article, it is widely believed that people can hold 7 ± 2 units of informaiton in working memory. The metric Miller adopted, however, was the threshold, defined as the point halfway between chance and perfect performance. If one is interested in the number of units people can *accurately* hold in working memory (i.e., without error), then the correct value is, as Mandler (1975) pointed out, 5 ± 2.

the Traveller model lacks a theory of how this importance comes about. And, although Navigator has an algorithm to measure subjective importance, it amounts to a simple count of how often the landmark is seen. However, even as simple an extension as counting frequencies represents an important step. The Navigator system uses such information at a higher level to simplify its scene comparison algorithms. Thus, the Navigator system is an example of how a theory of landmark representation can have a major impact upon higher level structures, something which the majority of cognitive mapping models cannot claim.

The better developed the theory about landmarks, the better developed the overall cognitive map theory is likely to be. This is because some of the properties of landmarks, such as the properties listed previously can have a significant impact on how the cognitive structure might be organized. The representational theory of landmarks that is assumed in PLAN is based upon Hebb's cell assembly (Hebb, 1949; Kaplan et al., 1991). Making the basic unit a cell assembly has a specific impact on the type of cognitive structure that it can participate in, namely, that it will be an associative network. Context in such a scheme can be handled through variably weighted links between the landmarks which, as it will turn out, is a natural solution to the path selection problem. By contrast, a system with a simplier theory of landmarks is less constrained in the types of higher level structures that it can build because landmarks can be treated essentially as tokens. As we shall see, many of these systems do model cognitive structure in manner very similar to our own, but the differences, however subtle, will be important.

## Topological Routes

In the constraint section of this article we proposed that humans actually develop two types of route map representations. One is essentially a topological structure consisting of landmarks, and the other is more directional and codes the relative spatial relationships between landmarks. In both cases, at the route level, these structures will reflect a relatively direct assimilation of experience. Our research group has already proposed a model that encodes the topological structure of landmarks, which we review in this section; in the next section, we propose a new model to encode directional relationships.

### Topological Networks of Landmarks: NAPS

In discussing landmarks we noted that landmarks are likely to be represented in an associative network. It turns out that a simple way in which to encode a topological system is into a network where the nodes represent landmarks, and directed links between them represent spatial proximity. In such a network paths could be extracted by following the links from one landmark to the next. The prototype representations of landmarks described in the

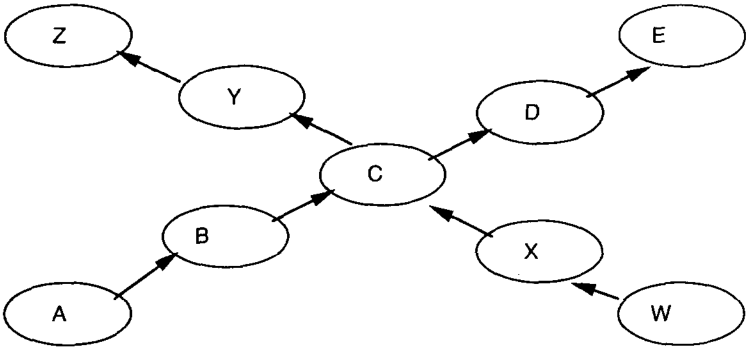


**Figure 1.** Traversing the paths A–B–C–D–E and W–X–C–Y–Z leads to the creation of the corresponding links. Because both paths intersect at C, the novel paths A–B–C–Y–Z and W–X–C–D–E can be extracted from the network.

previous section serve as the natural basis for such a representation. The contextual links between landmarks are exactly the links that code order and proximity information. A system which has been built upon these principles is called NAPS Network Activity Passing System (NAPS; Levenick, 1985, 1991). NAPS will serve as the basis for this portion of PLAN. What follows is a summary of some of the critical features of NAPS.

NAPS works by building sequences of landmarks when traversing an environment. NAPS is an explicitly connectionist model implemented as a network of nodes, corresponding to landmarks, and connections between nodes, representing the system's ability to go between two proximate landmarks. In such a system all of the information is stored locally; the only other landmarks that are connected to a landmark are those that can be seen from the first landmark. To traverse between distal landmarks requires an intervening sequence of connected landmarks. Such a structure is adaptively efficient for a number of reasons. Because the information is coded locally, there is little possibility for error; it is only necessary to recognize the individual landmarks. Such a structure can also be created with little experience, because all that is required is experiencing landmarks in sequences. Therefore, although such a topological encoding of the environment only contains a portion of the full spatial spectrum of information available, it does so quickly and relatively efficiently.

Compactness is achieved in an associative network by allowing individual sequences to overlap (see Figure 1). The intersection of sequences within a stored representation represents a major step beyond mere sequence towards cognition (Kaplan et al., 1990). A single sequence represents little more than a rote, repeatable act. Intersecting sequences, however, open up the possibility for the integration of solutions in a novel fashion. A simple example can be seen in Figure 1 where the sequence A–B–C–D–E can be switched to the novel sequence A–B–D–Y–Z after the sequence W–X–C–Y–Z is learned. In a cognitive map this means that one does not have to travel explicitly between two points to know how to get between them. It is worth noting at this point, that the problem of recognizing that sequences have common landmarks is not trivial. It requires the property that the object recognition system can identify landmarks from a variety of orientations, some of which are novel. This is one of the reasons that the type of solution to the landmark identification problem is so crucial to what kind of structure can be built at the path selection level.

The richness of the network structure does have a cost; route extraction is more challenging than in systems where routes are explicitly stored. The network structure also opens up the possibility of confusion because of the number of choices available at every landmark. Selecting the proper path from a given starting point becomes more difficult as more and more potential paths are added.

Levenick's (1991) solution to the problem is based upon a variation of spreading activation searches used in semantic nets and in some types of associative neural networks. The basis for path selection in an associative network consists of activating the representations for both the starting point and the goal location. As is generally the case in connectionist systems, activity is propagated from node to node where nodes in this case correspond to landmarks. Therefore, the activity of the start and goal landmarks will spread out from each other in something resembling a breadth-first search. Eventually the activity waves will coalesce (assuming the two points are connected by a path) at some intermediate point. This point then becomes a subgoal to be reached along the way. This process can be repeated using subgoals as start points and goals until a complete path is extracted. Conceptually, this type of selection process is equivalent to trying all possible paths out of the start state and all possible paths into the goal (later we will see that such a network is actually more sophisticated than this). When the two sets intersect, a candidate path has been found.

***Control Mechanisms.*** Levenick (1991) actually found that a strict spreading activation system is not viable because mechanisms are needed to control activity. For example, once a subgoal has been selected and the search process is repeated between the start and the new subgoal, some kind of inhibition is necessary to suppress the activity already existing in the network. A high level of performance was achieved that appears to be consistent with basic human cognitive mapping data by using various types of control mechanisms developed for systems of associative networks (O'Neill, 1990, 1991). The modifications that are of interest in this context concern variable strength connections and activity control mechanisms.

At this point, some of the differences between a system that is built using connectionist principles, such as PLAN, and more traditional symbolic systems starts to become clearer. The use of variably weighted links between nodes is an instructive example. A link between nodes represents the ability to get between two places. A simple representation, therefore, might have fixed links between two nodes in order to represent a path between the corresponding landmarks. However, using fixed links to connect adjacent landmarks yields an inadequate model of human wayfinding. Familiar routes are naturally easier to remember; routes that have only been traversed once or twice are going to be difficult to re-create. In terms of the learning rule used, this indicates that links between landmarks should have variable strengths. Familiar routes would be coded with high strength, whereas new routes would have low strengths; this is exactly what is done in PLAN. In terms of propagating activity during wayfinding, high-strength links will propagate a higher percentage of activity than low-strength links. Thus, nodes that are linked with high-strength links will tend to activate each other very quickly. One side effect of such coding is that in most cases familiar routes will be chosen over new ones because activity will propagate more quickly through the high-strength links. This is a conservative strategy, placing a high value on safety, with a strong emphasis on avoiding such hazardous outcomes as becoming confused or ending up in a dangerous place. This type of coding also results in the shortest path not necessarily being the path selected, because of the high premium placed upon the familiar. By contrast, symbolic systems with topological network structures, such as the Traveller (Leiser & Zilbershatz, 1989) and Mataric's (1990) system, spread activity in a uniform breadth-first search differentiating routes only by sequence length.

It is also instructive to note that at this level, PLAN is automatic rather than deliberative: There is no controller deciding which path to extract; the path arises from the structure of the network. Although it is true that under different circumstances the associative network of landmarks will generate different solutions, deliberative reasoning with such a structure requires higher level representations, some of which will be discussed later.

The other major control mechanisms used by NAPS, inhibition and fatigue, are commonly used in connectionist systems and in application are not directly relevant here (for more detail, see Levenick, 1991; Kaplan et al., 1991). The important pont is that theoretically meaningful constructs, such as fatigue, play fundamental and predictable roles in the management of activity.

*Hierarchy.* Most of the control mechanisms in NAPS are designed to dampen activity, making more differentiation possible. One problem that arises from such a design is that it is difficult to extract long paths from the network. The control mechanisms are designed to hold the activity in the network approximately constant (the variation and amount of activity is essentially equivalent to the activity of the $5 \pm 2$ landmarks that might be processed at any one time). The result is that, as the path between the start and goal states becomes longer, the average activity sustainable between them becomes less, eventually meaning that there are paths long enough such that no subgoal can be generated.

The solution to this problem was to add hierarchical elements into the system. As certain paths become well learned, a higher level representation of them is developed. Such a structure can serve as a compact representation of a path, which in turn could be a subpath on a longer path. Therefore, these structures are able to provide the support necessary for this type of associative network to extract long paths.

Although this form a hierarchy gives a topological structure, the capability for abstracting it is different than the kind of abstraction necessary for an overview of an entire environment. As a connectionist system, processing in PLAN correesponds to activity in its elements. Because this activity is passed from one landmark to the next, one at a time, there can be no coherent "view" of a route until it has been completely extracted. In an overview, on the other hand, an abstraction of an entire environment can be processed essentially simultaneously. The type of abstraction in a topological associative network is basically of the form of compacting known information. Although useful, it is not the same as an overview because an overview will require the ability to combine old information in new ways, such as figuring out a shortcut that has never been taken. Also, the hierarchy is still topological in nature and therefore does not represent certain, potentially useful, aspects of space.

### Comparison of NAPS to Other Systems

The defining characteristics of NAPS are that it is a topological model implemented in an associative network. Route knowledge is stored as variably weighted connections between landmarks. The major differences between this type of network and other systems that use topological information are where and how the knowledge is stored.

Tour (Kuipers, 1978; Kuipers & Levitt, 1988) maintains a topological model of learned environments, but stores route knowledge separately, in production rules. Tour has been criticized for this separation (Leiser & Zilbershatz, 1989; Miller, 1992) on the grounds that routes are independent, rigid wholes, for example, Kuipers admitted that this route knowledge is insufficient to find novel routes or shortcuts (Kuipers & Levitt, 1988). However, Tour has the capability to deal with these shortcomings to some degree by the use of its stored topological information.

By contrast, in Traveller (Leiser & Zilbershatz, 1989), as in NAPS, there is no separation between the topological network and the route knowledge. Like Tour, Traveller uses production rules to code routes, but Traveller more closely resembles a pure network model because routes are sequences of landmarks. As sequences of landmarks are traversed, production rules describing how to get from one landmark to the next are added. In this way a network of landmarks is built up. Subroutes of one sequence can be combined with subroutes of another sequence to determine completely new routes. The Traveller model still suffers from computational and storage problems, in part because it stores more than proximal information. Any time a path is traversed between two landmarks, the route knowledge is stored even when the two landmarks are widely separated in space.

Mataric's (1990) Toto robot uses a representation that is the most like NAPS: a network of landmarks. However, like the Traveller, Toto extracts routes in what amounts to parallel breadth-first search. In both of these systems this is seen as a desirable feature because searches will always return the shortest paths (at least in terms of the number of landmarks traversed). However, as we have already argued, this is probably not a sound adaptive strategy.

### Summary

We have built upon the basic structure laid out in the discussion of landmarks and used it as a reasonably efficient solution for basic wayfinding behavior. Despite the fact that it is essentially a pure topological model (versus some of the other systems discussed so far which incorporate other levels of space) and that the basic structure and learning rules of NAPS are fairly simple, it is capable of generating wayfinding behavior that is remarkably consistent with human data (O'Neill, 1990, 1991). Although its control mechanisms are complex, they are automatic, that is, they do not require any cognition to activate. NAPS, as described by Levenick (1991), does not explicitly address all of the issues of the structure of landmarks; nevertheless, it meets the necessary constraints. Levenick included control mechanisms that limited the amount of activity in the system to a level corresponding to $5 \pm 2$ objects. Context is handled through the connections between elements. The nodes in NAPS even have a certain amount of internal structure, controlling, for example, how long they can remain active.

### Locational Routes

One of the limitations of a topological system is the need for directional search. Even though humans are particularly adept at pattern recognition, such a search process is inefficient, and in a dangerous situation it might even prove to be fatal. After traversing a route a few times, a competent human does not have to search for each landmark along the way; on the contrary, knowing where to find landmarks becomes automatic. This implies that there is another level of space beyond, and perhaps separate from, the topological; we will call this level *directional* space. It must be pointed out that, unlike NAPS, most other models we discuss are not purely topological and do contain some notion of direction. However, the nature of these representations will be quite different, as we shall see.

The requirements for directional space are simple: When at one place, it would be useful to "know" the direction of the next place. Given the structure of the cognitive map developed so far, elements that are connected to each other are likely to be near each other. When one walks by A, one can next expect to see B if A and B are associatively linked. This means that it is not necessary to know the direction of any given landmark from any other; rather, it is sufficient only to know the direction of the landmarks that are associatively linked, and therefore close in space, to the current landmark, thus reducing the amount of directional knowledge necessary to a manageable level. Such a system would function in a manner very similar to an associative structure of landmarks except that instead of coding landmarks, the nodes in the directional system would code relative spatial information.

The problem faced in building a local directional representation, which we will call a *local map,* is fairly straightforward; the usefulness of a local map is based upon its ability to provide a relative change in orientation for any neighboring target landmark. Thus, when one is standing at the location corresponding to the local map, and when one desires to be facing a particular landmark, one should be able to use the local map to generate the relative change in orientation. Such a representation need not be exceptionally precise; once the orientation is fairly close to the desired value, the perceptual system can take over and use environmental feedback to increase precision as required. This is a major point of departure from the kind of directional information that is stored in most cognitive mapping systems. Directional information in cognitive maps is often based upon constructing geometric maps. Aside from the difficulty of such an undertaking (Brooks, 1985), such pure metric maps do not reflect the kinds of distortions that human cognitive maps are susceptible to (Passini, 1984). Other work, which acknowledges the difficulty and uncertainty involved in building a metric map, constructs maps that have been described as "rubber sheet" (Kuipers & Levitt, 1988) or "stretchy" (Brooks, 1985). We will argue that the human location system, which serves as the basis for how people acquire directional information, operates on a much more approximate level even than such rubber sheet maps, and that, given the powerful perceptual capabilities of humans, such approximations are sufficient to the task.

### Where Local Maps Occur

It might seem reasonable to suppose that local maps are constructed at landmarks. After all, the topological route structure codes journeys as being from landmark to landmark. In principle, this appears to be a sound idea, but in practice it has problems. First, one is rarely precisely at a landmark; rather, one is generally near them, using them as distal orienters. Landmarks are often large enough that being *at* one can mean a number of possible locations; a building has many sides, for example. Because local maps code visual information, it is the act of stopping and looking that is central to where they are constructed.

Local maps are useful when a new direction needs to be selected. At such times a person who is traveling is likely to pause and look around. This pause may be associated with contemplating which way to go, or it could be as a result of not knowing exactly which way the next landmark lies. In either case, the act of looking around from a single location is exactly what is necessary to create a local map. Therefore, one place that local maps will be created will be at choice points, such as at a fork in a road or a doorway. Generally, any journey will consist of traveling forward until some choice point is reached, picking a direction, and traveling forward again, repeating the process until the goal is reached. The choice points are where the local maps are needed because that is where a new direction may be selected.

Another place in which people are likely to pause and look around is when new information comes into view. When new information is afforded it pays, from an adaptive standpoint, to consider its consequences. In buildings, these places are typically doorways or intersections of halls; outside, they occur where a visual narrowing is followed by a visual opening, such as the entrance to a cave, an opening in a forest, or a pass through mountains or hills. In many cases these places also happen to be choice points.

### Local Maps

As we noted in the constraint section, for sighted humans the majority of spatial information comes through the *where* or *location* system. As such, the form of processing within the location should have a significant impact upon cognitive maps. One of the central hypotheses of this article is that the directional structures used in cognitive mapping directly reflect how information is processed within the location system.

***The Location System.*** Whereas the *what* or *contour* system is basically concerned with object recognition and therefore with the development of prototypes, the location system deals with the relationships between objects or, in extended space between landmarks. Because the systems have different functions, they use different information and different processing strategies. Aside from directional information, the location system is used in determining the size of an object and its relative distance. It is also sensitive to texture and surfaces. It should be noted that the location and size information is processed in the context of an implicit *picture plane* (Figure 2). The picture plane construct reflects the fact that at any given time the visual information

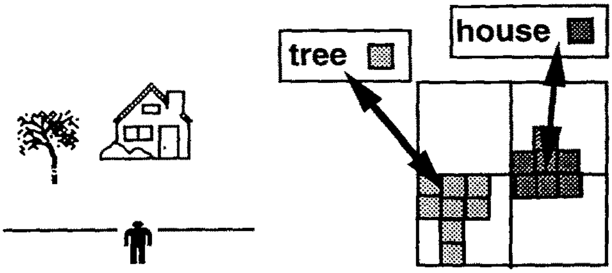


**Figure 2.** The scene as picture plane. Figure 2a depicts the observer looking at a scene consisting of a tree and a house; Figure 2b the observer's viewpoint is divided into four smaller regions. The objects are not detailed because of the mechanics of the location system. In principle, the scene could be divided into smaller regions.

being analyzed consists of a fixed scene. By treating the scene as a two-dimensional picture plane, one considers it to have the axes up-down and left-right. Within a scene an object's location will have specific coordinates along each axis.

Although the picture plane idea is two-dimensional, locational information is not restricted to two dimensions. In particular, depth information can be extracted through both textural and binocular cues. Nevertheless, the two-dimensional coordinate representation in itself is sufficient for storing directional information. It is not necessary to store more information because human perception, with environmental feedback, is fast enough to pick up more exact information as the system is used. Wayfinding only requires approximate direction. An exact direction may not be any more useful in any case because such a scheme relies upon being able to repeat the identical body and head positions.

***Location and the Eyes.*** The first step towards storing directional information is to determine where in the picture plane the landmarks are located. One of the primary tasks of the location system, referred to as segmentation, does just that. Segmenting a scene consists of dividing it into a small number of subregions, each corresponding roughly to an object or to an area without objects, that is, background. Lesperance (1990) implemented a connectionist segmentation algorithm that performs just such a task. Such an algorithm quickly yields the basic locations of objects within the picture plane.

Extracting the necessary directional information from such a scheme is straightforward. A scene can be represented by a grid, which corresponds to the picture plane. When the scene is segmented the objects or landmarks found can be linked to the corresponding grid cell. Use of the structure is also simple; when viewing the same scene the object can activate the appropriate cell through its connections and its relative location will be known automatically.

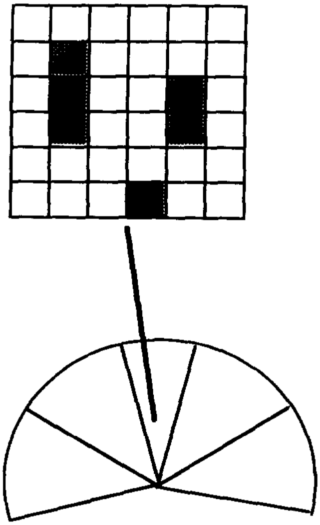
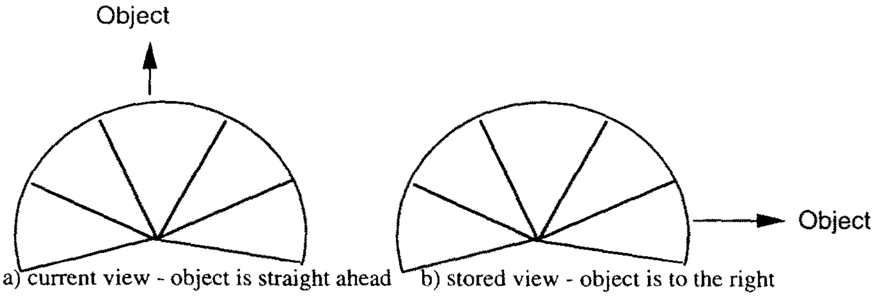


**Figure 3.** Each "slice" of the "pie" corresponds to a head position. The attached grid represents the scene that is in view with the head at that position. In this case the attached scene corresponds to what can be seen by looking straight ahead.

The link between the location system and eye movements has been supported by researchers studying the posterior parietal area of the brain. This research indicates both that there are regions whose receptive fields are retinotopic and that their responsiveness is modulated directly by eye position (Andersen & Zipser, 1990; Bushnell, Goldberg, & Robinson, 1981).

***Location and the Head.*** Shemyakin (1962) noted that there are three basic tools for orientation: the eyes, the head, and the body. The grid structure captures the method in which the eyes are useful for orientation within a scene. In turn, the head is useful for orienting between scenes; by turning one's head from side to side, a number of distinct scenes can be viewed. A simple way to organize scenes within a location is by storing them according to the relative position of the head. For example, one scene might correspond to what can be seen with the head turned to the left, another with the head turned to the right, and a third with the head looking straight ahead. A structure that captures this would be pielike, with each "slice" of the pie corresponding to an approximate head position and the scene relevant to that position (see Figure 3).

The pie is not complete because most heads have only about a 180° range of motion. The enormous biases people have to 90° and 45° angles suggest that the structure may be divided up into five parts; one straight ahead, two at 45° angles, and two to each side. We leave the exact specifications as an open research question.

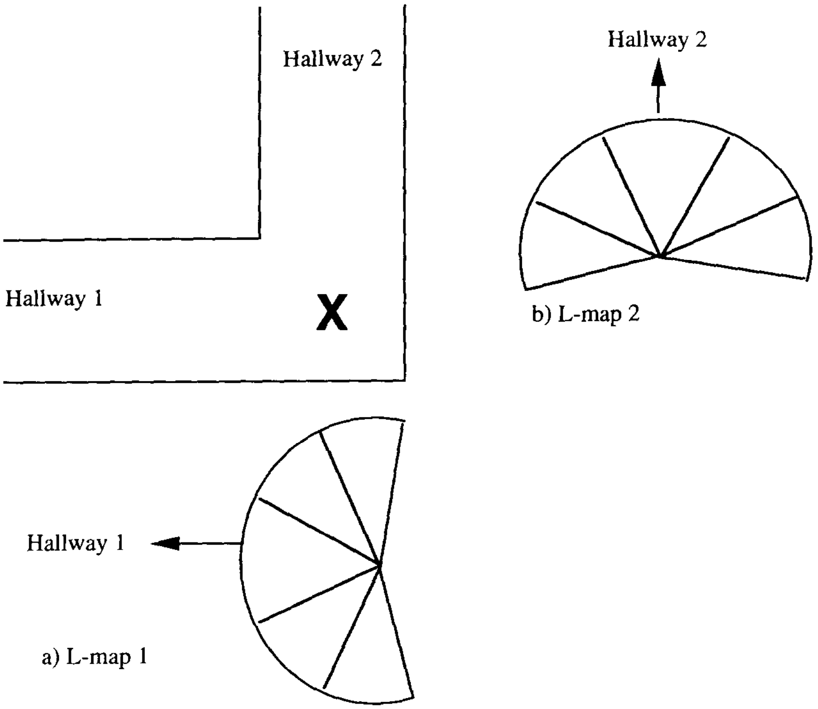


a) current view - object is straight ahead b) stored view - object is to the right

**Figure 4.** In the current view the observer is looking directly at an object. In the stored view, however, the object is directly to the right. This means that there is a 90° discrepancy between the two views. To see the normalized view, the observer should turn 90° to his or her right.

***Location and the Body.*** There is a difficulty with the head orientation scheme; it relies upon a fixed body orientation. The stored head positions are meaningless unless the current body position matches the implicit stored body position. For the representation to be useful, a normalized viewpoint is needed in order to ensure the correct body orientation. Such a viewpoint cannot rely on absolute directions because they are not typically known. If, however, one views a journey as a sequence of landmarks, then a solution arises quite naturally. The location that one has just passed can play the necessary orientation function. Assuming that one took a relatively straight route to arrive at the new location, the previous location will now be directly behind the new field of view. This location can serve as the orienter to the new representation; it is the point that should be directly behind one when the current view corresponds to the normalized viewpoint. In the future, when the location is reached, the structure can be oriented relative to this spot or to other stored landmarks. In fact, any two landmarks contained in the representation are enough to orient it with regard to the current view once the structures are in place. For example, in coming up to the spot of the representation, a landmark might be sighted. This landmark will have a normalized orientation within the representation. This orientation can then be compared against the current orientation to get a relative orientation. For example, if the landmark sighted is straight ahead in the current view, but at a 45° angle in the normalized view, then the current view corresponds to a view with the head at a 45°C angle from the normalized view (see Figure 4).

It is this combination of the grid and the pie structures oriented by the previous location that we are calling a local map. The name signifies that it is anchored at a particular spatial location and that the only stored information in a local map can be directly seen from that location.



b) L-map 2

a) L-map 1

**Figure 5.** Some locations potentially contain more than one local map (L-Map). Here, at the junction of two hallways there are two L-Maps, each facing down a different hallway, but both located at approximately the large X.

It is tempting to create a whole circle rather than just a semicircle in order to capture the full range of the directional information in one structure. However, this neither reflects the way in which people use their visual systems, nor is it easily constructed. There is a strong frontal bias to vision and to locomotion; local maps reflect those biases as evidence suggests (Shemyakin, 1962) that they should. The local maps arise naturally from the way the visual system and locomotion tend to work and are sufficient in most cases for orientation in wayfinding.

This does not exclude the possibility of representing the entire 360° at a single point; multiple local maps could be created at a single point. It does, however, entail a cost in switching between them. It also implies that for locations where there is no need for additional representations, only a single one will be coded. It seems likely that there would be a local map at each location corresponding to the various orientations one has when one arrives at that location along different paths (see Figure 5).

As was the case within the NAPS subsystem, connections within the local maps are not absolute, but rather reflect experience and the uncertainty involved in real environments. When arriving at a new location it is very difficult to determine which landmarks are permanent and which ones are transient. It is, therefore, far more adaptive to connect objects to representations with varying strengths, just as landmarks are connected together with varying strengths in NAPS. A stronger connection would represent a higher likelihood of the object being in a location, whereas a lower strength would represent a lesser probability. An object that is only seen once in such a representation would eventually lose its connection to the spatial representation. The connections in such a scheme more accurately reflect experience than would fixed connections.

## Networks of Local Maps

Local maps, like the landmarks in NAPS, serve as the basis for a route structure. There is no reason that the two representations should not be similar in structure, and indeed, our consistency principle along with the fact that our entire representation is connectionist both strongly suggest that the local maps are nodes in an associative network, which we will call an *R-Net*. There are two important implications of such a structure.

The first has to do with the problem of knowing which local map should be activated at any given time. If the local maps were stored with no structure, then activation of the correct one would require searching through all of them and choosing the best candidate. In an R-Net, on the other hand, there is no such problem. When one is at any given place and the corresponding local map is active, it automatically serves to facilitate the activation of the local map corresponding to the next place in the journey. This, of course, is a central contribution of an associative network. Search (in the sense of searching a large database of potential local maps) is not necessary; the correct local map will be activated by a combination of the predictive facilitation of the network structure on the one hand, and environmental feedback on the other. Environmental feedback is necessary because one local map may be associatively connected to several local maps and so will tend to activate each of them. By providing additional support to the correct local map as it is reached, environmental feedback enables that local map to dominate the competing maps.

The second implication was briefly discussed in the constraint section, namely, that there are two types of route structures available for wayfinding. This has several advantages. First, it provides a redundant system should there be a problem in the associative network of landmarks. Second, it is quite possible that the two systems working in conjunction with each other can operate faster and more efficiently, as appears to be the case with the two parts of the visual system (Rueckl et al., 1988). However, lacking simulation results, this must be considered speculative. Our conjecture is that for most people, R-Nets eventually supplant the associate network of landmarks as the primary structure for use in wayfinding. There are several reasons why this might be true. First, R-Nets more naturally reflect the experience of a journey. The places coded by local maps correspond to the places that one actually experiences while one is rarely, truly, at a landmark. Also, because R-Nets are richer in spatial information, they encode more of the information necessary to make a journey; in particular, they code which direction to take. Finally, as we shall see in our discussion of survey maps, R-Nets serve as the basis for the survey representation in PLAN.

### Comparison of Local Maps in PLAN to Other Systems

The idea of building metric-type information through a collection of stored views is not new. However, PLAN differs from previous implementations in what is stored, where it is stored, and the form of the storage.

A local map in PLAN stores scenes that link objects to their approximate locations. Other models generally store the entire scene. Tour, for example, stores "views" (which actually can include other sensory data) representing the traveler's sensory input at a given instant. Although Kuipers and Levitt (1988) claimed that these views can be abstracted and incomplete, it is not clear how this abstraction is achieved. There are two ways in which information is abstracted by local maps in PLAN. First, a scene will be divided into the $5 \pm 2$ dominant objects or regions that the location system identifies; only these will be processed. Also, for storage purposes, the objects stored with a scene will be linked to the scene with variably weighted links. This means that only salient objects will be coded with strong links, whereas other objects, such as things that are only seen once, will eventually fade from the representation. In this way PLAN is most like Navigator (Gopal et al., 1989; Gopal & Smith, 1990) which has salience measures to determine whether or not to include landmarks in its stored scenes.

Local maps are created in PLAN in response to basic environmental triggers such as at choice points or when new landmarks come into view. This strategy has the most in common with the Navigator system (Gopal et al., 1989; Gopal & Smith, 1990) and Mataric's (1990) work on Toto. Whereas Toto does not explicitly store scenes, it is sensitive to those places where new landmarks can be sensed. Because the landmarks are stored in a topological network, awareness of one landmark can lead to the expectation of another. Navigator stores scenes at decision points. Because Navigator is used to model navigation within a city, these places typically occur at street intersections. Other strategies include storing scenes at regular intervals (Asada, Yasuhito, & Tsuji, 1988), which has obvious drawbacks in terms of storage costs, and storing views when they are particularly distinctive as is done in Tour. Such a strategy makes sense if the collection of views is unstructured because it is necessary to uniquely identify the current view with regard to all possible views (Tour and NX actually create localized collections of views and therefore only require that a particular view be distinctive within a local region). However, if the scenes are stored in a network as in PLAN, then they not be distinct because even if two scenes contain essentially the same information, the appropriate local map will be activated by its connection to the previously active local map. Because systems like Tour do not store scenes in a network, the problem of matching what they are sensing to what is stored is significant enough that perceptual distinctiveness is perhaps the only reasonable strategy. The problem with a networked structure, on the other hand, is that is places a heavy burden upon the perceptual system to recognize that a new location is in fact a place where a local map has already been created. However, an unstructured collection of scenes essentially faces this problem with each location it reaches.

Finally, local maps in PLAN rely upon body, head, and eye positions for orientation. Other systems such as Tour and Qualnav (Kuipers & Levitt, 1988) attempt to fit the local views into a larger absolute space, whereas Navigator (Gopal et al., 1989) essentially sidesteps the issue by only working in city environments where turns are in 90° increments, and the streets provide a kind of coordinate system. At the level of the local map, PLAN makes no attempt to fit the local information into a larger picture, except when distal landmarks are in view; this is left to the higher level representations (to be discussed in the next section). Rather than trying to resolve information between two or more coordinate systems, local maps are anchored by neighboring landmarks and contain only approximate headings. The complementary function of the perceptual system greatly reduces the precision required in storage and computation. In PLAN, to head off in a new direction, the system would require a new facing (in 45° increments only), and within that facing, an approximate eye position. Once turned that way it can sight its target, get a more precise heading, and move on its way. By contrast, a system such as Tour would obtain a substantially more precise initial heading. This precision is only useful, however, if the currently calculated orientation is accurate, the stored heading is accurate, and an error-free turn can be executed. If there are errors, then the system will have to rely upon perceptual feedback to make corrections anyway. Because such errors are likely to occur, and because the human perceptual system is so powerful, there is little cost (and substantial advantage) in storing approximate information.

### Summary

The network of local maps both reflects and compliments the NAPS subsystem. The basic structure is the same in each case: an associative network in which the "nodes" are prototypes (see Table 1). The difference in struc-

TABLE 1
Two Network Structures for Route Extraction in PLAN

| Route Type | Basic Unit | Larger Structure |
|---|---|---|
| Landmark | Prototype | NAPS |
| Directional | Local Map | R-Net |

ture, and therefore in functionality, comes from what the prototypes are in each case. In the original associative network of landmarks the prototypes are objects in the world, whereas in the network of local maps the nodes consist of collections of scenes anchored at a specific viewpoint. Despite the similarity in basic structure, the character of each network is complimentary to the other even though each is capable of functioning independently to some degree.

As discussed earlier, the associative network of landmarks is a topological structure. Whereas such a structure is able to capture a great deal of spatial information—perhaps even the bulk of the useful information—other spatial information, particularly directional information, cannot be easily represented in such a format. The human visual system provides a wealth of directional information, however, and the local map structures capture the essence of this information in a structure that is economical and effective.

The critical concept in the development of local maps is the anchoring idea. There is no attempt to construct a single, objective, spatial representation, which would require integrating information across a large number of scenes. Instead, individual representations are constructed at critical locations, with the implicit assumption that these locations will tend to be visited on ensuing trips through the region. The end result is a structure that is relatively simple to construct and one that reflects the traveling and searching process rather than just being an objective structure. Such a strategy is common to a number of other systems (Kuipers, 1978; Leiser & Zilbershatz, 1989; Mataric, 1990). An objective structure would have two major disadvantages in this context: It would be enormously complex to construct, and it would have only a tenuous link to the process of moving through a space.

### Survey Maps

Neither the associative network of landmarks, the local maps, or even the R-Nets afford the ability to apply spatial operations directly to landmarks separated by great distances. For example, determining the direction of a landmark that is even moderately far away would be extremely difficult using just these representations. All three structures are maximally informative with regard to things that are close to the current location. But people also know about the relationships between landmarks that are not close to each other. Such information can be used in spatial reasoning, such as when determining if there is a shorter route between two points than the standard path, in facilitating the search in either of the route structures, or even in providing the capability for hierarchical planning.

In discussing the creation of local maps it was pointed out that some places in the environment are more obvious locations for local maps than others. This suggests that there may be places where it is natural to pause, look around, and reflect on one's choices. Such reflection may allow one to extend the corresponding local map beyond what can be immediately seen by using the cognitive map's predictive capacity, thereby building larger scale directional representations, which we will all *regional maps*.

### Gateways

The two major qualifications for creating a local map were that a choice point had been reached and that new landmarks could be seen. We shall refer to a place that meets both of these qualifications as a *gateway* in honor of Christopher Alexander's design construct (Alexander, Ishikawa, & Silverstein, 1977) which is strikingly similar in description to the functioning of our own representation. In buildings, these are typically doorways; outside, they occur where a visual narrowing is followed by a visual opening, such as the entrance to a cave, an opening in a forest, or a pass through mountains or hills. Therefore, a gateway occurs where there is at least a partial visual separation between two neighboring areas and the gateway itself is a visual opening to a previously obscured area. At such a place, one has the option of entering the new area or staying in the previous area.

As places that tend to separate different areas of space, gateways are a natural place to begin building up a higher level of directional space. The first level of directional space is local: A local map is defined by the area that can be directly seen. A gateway, on the other hand, is often an entrance to a larger space. Such a space can be defined as the area between gateways. For example, in a building one is in a particular room until one passes through a door to another room, or one is outside until one passes through a door leading inside.

### Regions

Regions are defined by visual barriers and gateways; examples include the walls and doors of a building, the hills and passes between them in valleys, and the trees and paths into them in forests. Not only are gateways the starting points for building a spatial representation of a region, but they tend to be the places that are visited the most often. A building cannot be entered, for example, without going through the entrance. In the literature, places that are strongly represented and vital to the organization of regions have been called *nodes* by Lynch (1960), *centroids* in Traveller (Leiser & Zilbershatz, 1989), *anchor points* by Golledge (1987), and are vital for the organization of Yeap's representation (1988). Of these, our representation has the most in common with Yeap's. Yeap divides space up into what he called absolute space representations (ASRs), which correspond, in a building environment, to individual rooms. However, while the gateway notion emphasizes the importance of transitions between spaces, Yeap constructed his local representations at the middle of each room. This was done to maximize the amount of information about the local environment that could be stored relative to a single location. Although Yeap's work has influenced our own, the difference in philosophies as to the locus of the critical points for storing information is telling. This difference ultimately leads to quite different kinds of systems.

Yeap (1988), like many researchers in AI and robotics, was interested in optimality. In this case, Yeap would have liked to have found a point in a room where the entire room (if possible) could be seen. If such a point could be found, then it would be possible to store all of the spatial information associated with the room at that single location. Given this as a goal, the best places to use tend to be in the center of the room in question. The gateway notion is built upon different principles. Among these are prediction and choice. Because gateways occur at transitions between spaces, they represent locations at which a choice will have to be made: to remain in one space, or to move into the next. Because gateways occur at choice points, they are places where people naturally pause; this pause allows scenes to be more carefully parsed and analyzed, and therefore makes them more likely to be represented and remembered. If an entire space cannot be seen from a gateway, then more local maps will be created as needed when traversing the space. Whereas such a system may not ultimately minimize the number of points where representations have to be created, it is nonetheless practical and, we would argue, more efficient because complex transformations are not needed. Because the structures on which PLAN is based reflect experience so closely, it is a simple matter to put them to practical use. However, structures designed explicitly for optimality of storage or other concerns may not be so practical because they will surely require extra processing; a structure based upon absolute Euclidian coordinates, for example, does not easily lend itself to providing eye locations or simple turns of the head.

Another difference between gateways, on the one hand, and nodes, centroids, and so on, on the other, is one of levels. Whereas gateways are rather low-level representations based upon visual information, the other types of points are higher level, playing a role akin to abstraction. A gateway, by contrast, is a vantage point, a concrete experienced place from which an array of visual information is available. One way of thinking about this difference is that in most representations one point will stand in for an entire region, as in the Traveller system, whereas gateways represent transitions between regions. It is a polarized or directional view of a room: a view from the choice point rather than an objective, optimal, already-there view. Therefore, it is our contention that gateways represent a different kind of representation; in the next subsection we will discuss a component of our representation that is much closer to the centroid concept.

The gateway concept is, in our view, pivotal to the creation of regional structures. Not only are gateways visited more often than other places, but they also provide an exit point. Such an exit point is important not only as a potential means of escaping a dangerous situation, but in exploring a new place, the exit point provides a place where one can return to the old, familiar environment.

The survey maps which we are proposing, called *regional maps*, have the same basic structure as the local maps. This has the advantage of consistency as well as of continuing the emphasis on vision, the primary mode of functioning for humans. There are times when distant objects can be seen in relationship to each other and such a view affords uniquely valuable information. An example of this would be when observing from a height, such as a hill. By looking from an oblique[3] viewpoint, landmarks do not obstruct each other; thus their relationships to each other can be observed simultaneously. Again, the gateway is an ideal location for such a viewpoint. The region is defined as being entirely on one side of the gateway so that the entire region can be "seen" from that vantage point. Of course, the scenes actually viewed from gateways aren't necessarily oblique, and a major issue, which will be dealt with in the next subsection, is how the equivalent of an oblique viewpoint can be constructed.

## Regional Maps

The R-Net structure gives rise to the development of regional maps. At any given time a local map will be in use. Locomoting around an environment will mean that local maps are activated and deactivated sequentially. As a familiar path is taken, each local map will begin to predictively activate the next local map in the sequence even before it is reached, for example, "around this corner I would expect to see. . . ." As familiarity with the region grows, the predictions will become stronger, faster, and more accurate. In time, the predictive effects of association will begin to be taken into account in the current local map, particularly if one pauses, as one might, at a gateway. If one mentally runs through the next part of a journey, a larger structure containing not only the current local map, but parts of the neigh-

---

[3] *Oblique* is a term used by draftsmen and environmental designers to identify a viewing angle somewhere between an eye level view and a plan (or top) view. Geographers have determined that an angle of 30° is generally preferable in presenting map information, a tradition that harks back to medieval times when views of a city were often drawn from the perspective of a person on horseback.

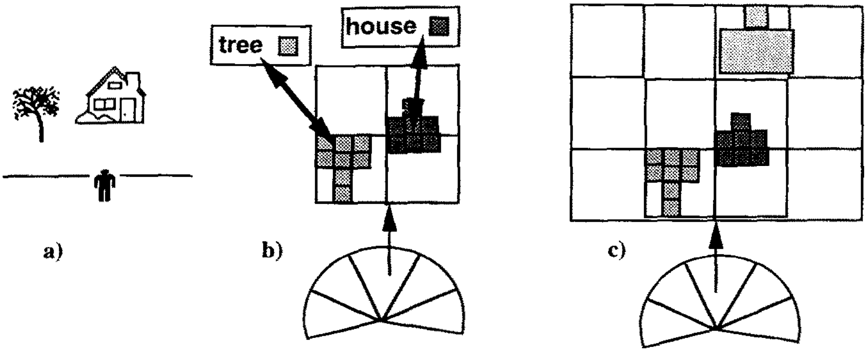


**Figure 6.** Figure 6a shows a person standing at a location; Figure 6b shows one "slice" of a local map for the location, and Figure 6c represents a possible regional map at the location. The local map has simply been expanded. In this case the new shaded figure represents a house directly behind the house in 6a. Its higher position in the regional map indicates that it is farther from the observer. Such a representation would be consistent if one were floating above the original viewpoint because one would then be able to see both houses directly, and the second house would indeed be behind the first with such a view.

boring local maps can be created. Such a representation would contain information beyond that which can be seen by taking advantage of the predictive power of a network structure. This approach is similar in spirit, though different in mechanics and result, to the way in which large-scale geometric representations are built in the NX robot (Kuipers & Byun, 1991).

At this point it is important to remember that the local map concept was based upon the idea that what should be stored is what can be seen from a particular vantage point. In the regional map case, however, information is stored that cannot be seen from the stored viewpoint. This could potentially present a problem in the use of the structure. For example, if one landmark is directly behind another, then they would occupy the same location in the visual field. Thus, when the landmark that was behind is added to the grid representation, it would occupy exactly the same grid square and therefore appears to be in exactly the same location. Fortunately, there is a simple solution to this problem. If something farther away is thought of as being farther *up* or *out* in the visual field, then more distant objects will be placed farther than previous objects on the periphery of the grid structure. Such placement represents a distortion of the true visual field at the location in question. However, the distortion can be resolved by considering the regional map to occur at a new point directly above the original point, corresponding to an oblique view of the mapped region. Taken further, this means that as the area covered by the regional map expands, the perceived *height* of the map will rise (see Figure 6). Thus, an oblique viewpoint emerges as one becomes more and more familiar with a large-scale space.

It might appear that regional maps are simply local maps that grow larger and larger as the environment becomes more familiar. If, however, local maps simply grew until they contained entire regions, they would contain an unmanageable number of landmarks. By contrast, if regional maps are conceptualized as abstractions of the information contained in the associative network of landmarks—the local maps and the R-Nets—then they should reduce the amount of information to its most important essence, not simply provide another organization of it. This loss of information does not constitute a handicap because if more information is needed it can be extracted from the lower level representations.

The reduction in information comes from the context in which regional maps are used. Regional maps function primarily in the planning process. A high-level plan of a trip may consist of only a few key landmarks even when the journey is long. Because it is the nature of connections to reflect experience, if certain landmarks are used repeatedly in plans then they will become part of the regional map. Landmarks that are rarely used will fade from the representation just as transient objects fade from the representation of a local map. The landmarks that will remain in a regional map will be those that are used over and over again in the context of the large-scale being represented. It is in this context that the landmarks stored in regional maps function similarly to centroids, nodes, and so on, A regional map of a city, for example, will have only a few landmarks. High-level plans formed using such a map would go from one of these landmarks to the next, exactly as might occur with the centroids of Traveller or in the plans of the taxi drivers in Pailhouse's study (1969). This strategy of solving problems by starting at a high level of abstraction and working towards more detailed analysis is found in numerous AI systems starting with GPS (Newell, Shaw, & Simon, 1960) and ABSTRIPS (Sacerdoti, 1974).

Regional maps function within the hierarchical structure of the larger system. Thus, regional maps can be used to generate larger regional maps. The process involved would be virtually the same as in generating regional maps from local maps.

### Using Regional Spatial Representations

There are two major advantages afforded by regional maps. First is the added ability to do hierarchical planning. Regional maps potentially can model just as large a space as NAPS but will contain more spatial information and far fewer landmarks. Even at the level of a city, a regional map will contain just a few landmarks. Plans formed at this level must be simple; this simplicity makes planning efficient and diminishes the chance for confusion. Each stage of the high-level plan can then be broken down into a smaller plan and the process repeated. Such a scheme is in accord with the well-known study of taxi drivers done by Pailhouse (1969). In that study, taxi drivers appeared to have divided up the city into smaller regions. When traveling to a new region, they first went to a standard point in that region before proceeding to their ultimate destination. At the high level, the city seems to consist just of a few regions and one particular point in each region. However, each region in turn has a more detailed representation. Such a framework is also useful in dealing with breakdowns in plan execution. A bridge that is out will not undermine the entire plan, but only one section of it.

Regional maps can also be useful in performing certain types of spatial reasoning. Because the spatial relationships of distal objects can be "seen," it is possible, to some degree, to determine whether a certain path is spatially efficient or whether it wanders too far in any direction.

### Comparison of Regional Maps to Other Representations

Regional maps provide PLAN with the ability to make simple visual abstractions of large-scale environments. Just as with the local maps, a significant feature of these abstractions is the environmental configurations that trigger their creation. These points, called gateways, are critical in that they determine the boundaries of the regions and, because the abstractions are visual, constrain what can be "seen" in any abstraction.

One obvious comparison is to the Traveller, which breaks down large environments into regions, each containing one centroid. These centroids are then used as focal points for long paths. Travel over long distances is generally viewed as from the current location to the nearest centroid, then to the centroid in the target region, and finally to the target location. This is quite similar to how planning would work in PLAN when using the regional maps. Because only a few places will be represented at this level, plans will tend to focus upon those points. One minor difference is that regions in PLAN will contain more than one such point (although this may be merely definitional because each region could be broken down into subregions surrounding each place in a regional map). More importantly, PLAN contains an explicit description of the circumstances in which these places and regions are formed, whereas in the Traveller their creation is not well specified. Indeed, the optimal locations for centroids are described in terms of distance from other centroids rather than as being responsive to the configuration of the environment.

In Tour, on the other hand, the regions are well defined. The edges defining a region are determined by the paths that have been previously taken. Essentially, a path defines a boundary between things to the left of the path and things to the right of the path. As boundaries intersect they will begin to define enclosed regions. Whereas the centroid idea came from Pailhouse's (1969) taxi driver study, the boundary idea is derived from work that indicates that many human cognitive maps contain "skeleton maps" of major streets. In this case, the frequently traveled streets define the skeleton or boundary structure in Tour. Regions in PLAN do arise out of use as they do in Tour, but whereas the defining characteristics in Tour arises out of motion (traveling on a path), in PLAN it arises through stopping and looking around. In PLAN, and in Traveller, the abstraction at the region level was from place to place. In Tour, with its emphasis on paths, the abstraction is from path to path. So a route-finding heuristic would be specified as going from the current location to a well-known path, taking that path, and then going from the end of that path to the goal location. There is also a familiar path bias in PLAN. It occurs, however, not at this level, but at the level of the route structures where familiar routes are coded with stronger links, and therefore are retrieved faster and more often, and may also be directly abstracted. Tour, on the other hand, does not have a global overview that can quickly yield the relationships between distant landmarks.

### Summary

Regional maps build upon the simple functionality of the where-to-look mechanisms. However, although they are structurally identical to local maps, they bring substantial extra power to the cognitive map. This power is due to the hierarchical capabilities inherent in regional maps. Regional maps are simple abstractions of large spaces; a regional map for a city might only contain five or so landmarks out of an extraordinary number of possible alternatives. Once again, the environment, in conjunction with the limited capacity of the system, has dictated the representation. It is the usefulness of a landmark that determines whether or not it will occur in a regional map, not its geometric location. The power of regional maps is a direct consequence of this simplicity. When more information is needed, one can always go down a level in the hierarchy.

## EVALUATING PLAN

One of the major themes of this article is that the individual pieces of the cognitive map should mesh together into a smoothly functioning whole. Now that all of the pieces are in place, it may be helpful to go through an example of a possible cognitive map within a specific environment. In doing so we will evaluate PLAN in light of the constraints put forth earlier. Following that, we will make general comparisons between PLAN and other models of cognitive mapping.

### An Extended Example

For this example we will consider a hypothetical world originally presented by Kaplan and Kaplan (1989) called "John's world." John has several routines that play a prominent part in the development of his cognitive

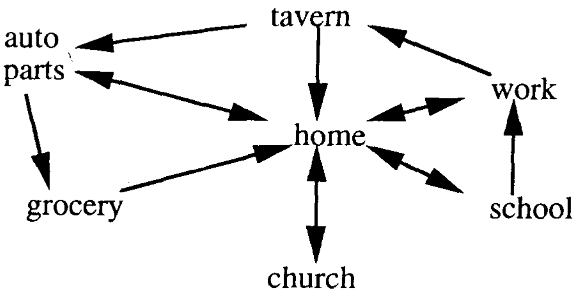


**Figure 7.** A network representation of John's world.

map. He goes directly to work five mornings a week, except in bad weather when he takes the kids to school even though this constitutes a detour from his customary route to work. On Saturdays, he often goes to the auto parts store, and on Sundays he sometimes goes to church. Sometimes after work John drops by the neighborhood tavern on his way home. On Fridays, John sometimes goes home from the tavern by the way of the auto parts store, to save having to go there on Saturday morning. Finally, John often goes from the auto parts store to the grocery store when his wife asks him to pick up a loaf of bread.

The major landmarks in John's cognitive map correspond to the common destinations in his routine (See Figure 7). These are the places that are the most meaningful and useful to John. John builds up representations of these landmarks that reflect his experience, for example, placing them in specific contexts with particular orientations.

As John travels around town he begins to know what to expect next during his travels. This is due to his associative network of landmarks, which has a kind of predictive quality. The nodes in this network correspond to the landmarks that John has learned. Contrast the amount of information stored in such a network with the total amount of information contained in John's neighborhood (see Figure 8). The network structure has the additional advantage that John can easily extract routes, such as from the grocery store to school, that he may never have experienced.

Concurrent with the development of his associative network, although slightly slower, will be the development of John's local maps. As John learns to navigate through the environment, he will come to learn the twists and turns of the familiar paths. When John first moved to his neighborhood, determining which landmark followed the next was such a challenging task that he had little attention left for determining spatial relationships. However, as John becomes more familiar with his environment, his underlying associative network becomes more efficient, freeing up capacity for the task of learning the directional relationships of the landmarks. It is at this

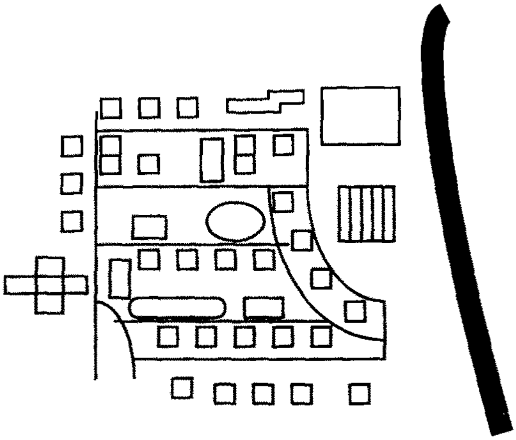


**Figure 8.** Aerial view of John's world.

stage, when the local maps first start to become solidly learned, that John can become proficient at getting around his environment; the internal management of his associative networks is already fairly sophisticated and with the local maps there is a reduced need to look around to figure out where he is and where he is going.

As John learns local maps, they too are structured into associative networks corresponding to R-Nets. In the early stages, the function of R-Nets is almost purely predictive: As John leaves the location of one local map, the R-Net structure is useful in readying the next local map for its use. Later on, when the R-Net becomes well learned, it serves as a redundant system for wayfinding, possibly even supplanting the associative network of landmarks. When functioning in such a mode, John will be working in a strikingly similar manner to the children described by Shemyakin (1962) who drew their neighborhood by taking an imaginary journey.

Increased facility with the use of the R-Nets will lead to the development of an environmental overview or regional map. In this case John is familiar enough with his environment to know the spatial relationships between landmarks even when they are visually separated. When using his regional map, John will have the feeling of viewing the scene as if from above, with the landmarks seen in oblique perspective (see Figure 9). John can then use this spatial information to perform efficient spatial operations. For example, in going from his home to the grocery store John can "see" in which direction it lies and that a reasonable path might be to go by way of the auto parts store, whereas a route that goes by way of the church would be quite inefficient (see Figure 10).

When John "sees" his neighborhood he is functioning in a manner that appears to be in accordance with survey maps as described by Shemyakin

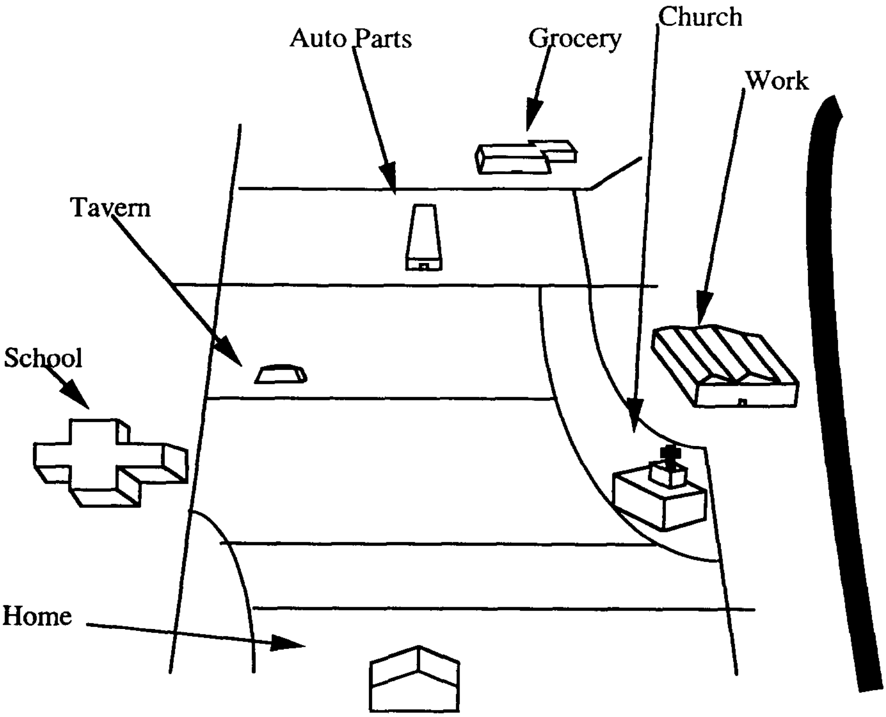


**Figure 9.** Perspective view of John's world from a point above and slightly behind John's house.

(1962). However, as we noted in our description of survey maps, such a conceptualization appears to violate Piaget's hierarchy (Piaget & Inhelder, 1967) which specifies that survey maps are objective. Indeed, the viewpoint notion, which serves as the basis for regional maps, still appears to be egocentric. However, the viewpoints of regional maps, being generalized, normalized, and synthetic, contain more information than a single view. The factors that lead to the creation of a regional map at a particular location are such that the resulting representation is *apparently* objective. Because the viewpoint can take in the entire region at one time and (through hierarchy) contains all possible objects and their relationships, the result is that it appears as if the viewpoint is objective and serves as an approximation thereof. It is also the case that because John is creating local maps (and regional maps) at multiple places in the environment, he can switch to the viewpoint most appropriate to a given task. Because he has this ability to use different representations of the same environment, the overall representation is not egocentric. The combination of the particular viewpoints used and the number of viewpoints available makes the objective frame of reference question moot.

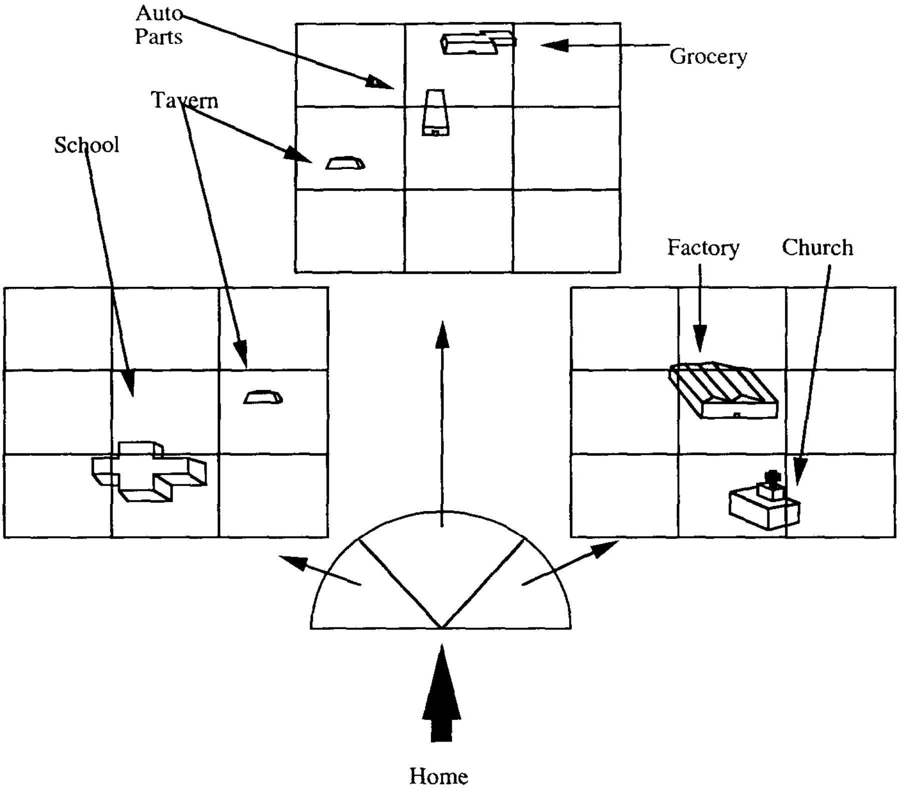


**Figure 10.** One "slice" of an R-Map anchored at John's home.

John's cognitive map is very much in line with our criteria. Such a cognitive map representation is attractive from a variety of standpoints. Though the entire process of learning the map is relatively slow, it is nevertheless useful even at an early stage. With each development comes increased functionality, and generally, a corresponding increase in the speed at which the map can be used. However, whereas most of the mechanisms are consistent from structure to structure, the individual pieces rely very little upon each other. Economy is emphasized through the development of the map; things are only added as they prove useful, and the structure that is stored is of low precision, with the resulting benefit that complexity is greatly reduced. As we have noted throughout, the basis for the directional structures in PLAN comes from research on the two visual systems in humans. It is because of this work that some of the minor differences between PLAN and developmental theories have arisen. In particular, we are proposing that there are two types of route structures rather than one, and that the structures which are called survey maps are not truly objective, but only appear to be. Another extension we are proposing to the developmental literature is to consider landmarks in light of the object recognition and categorization literature, particularly with regard to prototypes.

### Models of Cognitive Mapping

The models that we are comparing with PLAN are summarized in Table 2. Of these models, Qualnav (Kuipers & Levitt, 1988) and Mataric's (1990) model do not claim to be models of human cognitive mapping although they appear to derive a certain amount of their theory from such work. Most of these systems have commonalities with PLAN, and all of them have some type of topological structure. PLAN differs from all of these systems in that it is connectionist and builds global overviews of the environment.

At the landmark level, the Navigator (Gopal et al., 1989) system probably comes the closest to having a theoretical treatment as extensive as PLAN's. Other systems such as Tour (Kuipers, 1978), Qualnav (Kuipers & Levitt, 1988), and Mataric's (1990) decide that an object is a landmark based purely upon sensory distinctiveness. Both PLAN and Navigator acknowledge that this is the critical issue in landmark formation, but both also acknowledge that issues such as importance can play a significant role (the Traveller model does as well, but does not have an explicit theory as to how such factors work). Typically, this importance would reflect how often a place is used, for example. PLAN goes beyond even Navigator both by addressing issues such as context and how landmarks might be processed, and by putting landmarks within the framework of categorization.

The next stage in cognitive map development is the route map level. At this level we have proposed that there are actually two types of route extraction mechanisms. The first is based upon a topological network of landmarks and could be called the traditional route extraction method. Such a method is used by Mataric's (1990) system and Traveller (Leiser & Zilbershatz, 1989) as well as PLAN. Such a method of wayfinding is natural because landmarks are what people learn first in new environments. However, such a structure does not accurately reflect a typical journey because travel does not naturally occur as landmark to landmark; one is often near landmarks, but not usually *at* them. The second type of route map is based upon networks of places, where places correspond to the locations that one would actually pass through in a journey. Examples of such places include decision points in Navigator (Gopal et al., 1989), enclosed regions in Yeap's (1988) system, and gateways (Alexander et al., 1977), which may be decision points or places where new landmarks can be seen, in PLAN. The network of places can take advantage of those distinctive locations that occur in wayfinding and use them for a mode of route storage that more naturally reflects the travel experience. Gateways in PLAN are more general than the decision

TABLE 2
Comparison of Cognitive Map Models

| | Routes | | Direction | Survey Maps | |
|---|---|---|---|---|---|
| System | Topological network of landmarks | Other route identification methods | Directional structure | Basic survey map mechanism | Presence of global overview |
| PLAN | Yes | Network of scenes | Qualitative | Locational mechanism with scene abstraction | Yes |
| Mataric (1990) | Yes | None | Metric | None | No |
| Navigator (Gopal, Klatzky, & Smith, 1989) | No | Network of decision points | City block | None | No |
| Qualnav (Kuipers & Levitt, 1988) | No | Computation based upon landmark triangulation | Metric | Topological map of viewframe | No |
| Tour (NX; Kuipers, 1978) | Yes | Stored rules | Metric | Regions defined by paths | No (NX does contain a global metric map) |
| Traveller (Leiser & Zilbershatz, 1989) | Yes | None | None | Regions represented by centroids | No |
| Yeap (1988) | No | Network of regions | Metric | Each place map considered a survey map | No |

points of Navigator because they encompass places where one is likely to pause, and more specific than Yeap's regions, which can cover a fairly significant area. Only Tour (Kuipers, 1978) and PLAN have the capability for both types of route extraction. Tour can extract routes through a topological network just as PLAN can, and Tour's basic route structure is based upon capturing the structure of a journey just as in PLAN. However, while PLAN schematizes journeys into networks of places, Tour stores rules about getting from place to place. The result is that paths are generally treated as inseparable wholes in Tour; the places along the way are not significant beyond being a part of a path. Tour does have, however, the capability to combine pieces of different paths when necessary. PLAN, by contrast, retains the distinctness of the landmark representations that define the paths, permitting intersecting paths and, ultimately, the formation of a flexible network structure. Another difference between the two systems is that Tour does not make the developmental distinctions that PLAN does, simply positing that the path structure will generally be used and the topological network will be used when there is no known path.

A survey map is a compact representation of a large-scale environment. The survey maps of PLAN, called R-Maps, are visual in nature and correspond to what one might see if one were looking out at the environment from above. However, the survey maps of PLAN are still compact because they contain only a fraction of the possible information that could be stored. This schematized global overview is unique in the cognitive mapping literature. Navigator, Tour, and Qualnav store views, but do not abstract them beyond the local level. Traveller does have a compact large-scale representation of a sort, but it could not be called an overview, and Traveller lacks a theory as to how these points are developed. Both Yeap and Tour do have explicit theories about regions, but neither has a theory for developing them into an abstract hierarchy. The structure of R-Maps lends itself to making predictions about human functioning. First, because the R-Maps correspond to particular locations with particular orientations, which we call gateways, it should be the case that there are views of familiar environments that people prefer. For example, someone who lives on the north side of a city might prefer a map which has south as up rather than north. Porteous (1971) showed that this is the case and that people can in fact be confused by a map with north as up. More recently, Warren and Scott (1993) showed that people prefer to align maps with the environment it represents and that their performance on wayfinding tasks is better when such alignment is done. Another prediction, for which we only have anecdotal evidence to support, would be that a map which is anchored at such a point and provides an oblique perspective should be preferred over a standard cartographic map with a plan or overhead view.

A final point of comparison between PLAN and these other systems could be labeled as the style of the representation. The representations of PLAN tend to be schematic: Less is stored rather than more. The problem with such an approach is that is relies upon being able to differentiate between what is important and what is not. Traditionally, what has been important in cognitive mapping has been landmarks: Landmarks are the most obvious salient features of the environment. One of the advances of cognitive map theory has been the identification of what else in the environment is important. Lynch (1960), for example, identified junctions, where paths come together, as being important places. Navigator, among other systems, extended this idea by identifying decision points as being particularly salient. The Traveller system takes this notion even further by defining a kind of super junction, called a centroid, which is a location of special importance. PLAN attempts to provide a framework to explain what it is that makes these and other locations so important. One such issue is the structure of the environment. Some places are important because they open up a new vista of information. Another issue is usage. Some places are important because they are visited time and time again. As in the Navigator system, some places are important because they are at decision points. PLAN also takes these concepts a step further by defining a gateway to be a place that represents the integration of all three of these factors.

## CONCLUSION

One of the fundamental ideas that differentiates PLAN from other cognitive mapping systems is that it is a "heads-up" or scene-based representation. The stored views in PLAN are not from an aerial perspective as in a typical map, but reflect what an observer sees through different head positions at a single location. In this, as in other matters, one of the key tenets of PLAN is that storage reflects experience. Whereas an overhead representation might cover an entire area, it may be necessary for multiple scenes to provide full coverage of an environment in PLAN. However, these scenes will correspond to locations that one is likely to be in. Furthermore, it turns out to be a relatively simple matter to extract useful information from the stored scene and to use it directly. For example, in PLAN one can easily extract head and eye locations for the positions of neighboring landmarks. By constructing a "semantically transparent" representation, the job of putting the representation to use is considerably simpler (Clark, 1989; Smolensky, 1988). In an overhead representation one must translate from the current perspective to the overhead perspective and back, a potentially confusing and complex task. Although PLAN is not the only scene-based system [see, e.g., Kuipers & Levitt, 1988 Tour & Qualnav], the scene information stored in PLAN is schematic; the perceptual system is fast enough and flexible enough to take an approximate direction and convert it to an exact location.

Central to the success of the heads-up perspective is the idea that certain locations in the environment are particularly important because they are visited frequently. In PLAN, these locations are called gateways. Again, the gateway notion comes about from taking an adaptive approach to building structure. Gateways are natural places for building scene-based representations because it is at these locations that a new vista of visual information is available. Organisms also need to know about gateways because they represent choice points and escape opportunities. Finally, gateways are important simply because they are visited so often: It is not possible to enter a new space without passing through a gateway of some type.

Many buildings intrinsically incorporate the principles of gateways into their design. Maps of the building are usually located right next to the entrances. Additional floor maps might be included next to the elevators on each floor, and pointers to rooms are often put up at hallway intersections. All of these locations are gateways because they all occur at places where new visual information becomes available, from the whole new scene when stepping out of an elevator, to the view down intersecting corridors. Such a mapping scheme is so natural in the context of buildings that it is hard to imagine any alternatives. The building environment is such that all places are not created equal; places such as doorways and the intersection of corridors are especially important because they represent decision points. We argue that the same can be said for the outdoor environment with cave mouths, mountain passes, entrances to forests, rivers, and so on. This is another difference between PLAN and other cognitive mapping schemes: The structure of the environment plays a particular and central role in the construction of the representation at multiple levels. Although environmental distinctiveness is an important part of most cognitive mapping systems in terms of landmark recognition, at higher levels the environmental relationship tends to become fuzzier. Traveller, for example, has a region structure, but no natural method for determining regional boundaries. Tour, on the other hand, uses paths as boundaries. PLAN is the only system in which the regional structure is an analog of the physical and visual structure of the environment.

PLAN synthesizes elements of Gibson's (1979) perceptual theories with an internal representation concept. Gibson contended that humans do not have such internal representations, that it is not necessary to store everything that is in the environment because the environment itself will provide much of the information that one needs. A modern version of this idea, though not specifically intended as a model of human cognition, can be found in "behavior-based" models (for reviews, see Brooks, 1991; Maes, 1993). For example, it is not necessary to store every feature of a tree because when the tree is next seen all of its features will be available; it is only critical to tune the system to respond uniquely to the tree. This could be called an environmentally centered approach because it builds on the idea that the environment is rich enough to provide all needed information, and furthermore, the internal representations are extremely sensitive to the environment. A more traditional information-processing view, on the other hand, is the knowledge-based approach. The goal of many such systems is to capture as much knowledge about the task domain as possible. Although this approach, when applied to cognitive mapping, does store environmental information, it does not necessarily do so in a transparent fashion. Once an environment has been learned, all planning and reasoning can be done completely internally and exhaustively. By contrast, a system like PLAN only stores a fraction of the available information internally and relies upon the perceptual system to fill in gaps when plans are put into action. An information intensive approach is perhaps sensible for robots with limited perceptual skills, and for computer simulations where there is no real environment. Humans, however, bring an impressive array of perceptual skills to bear in environments that are too complex to learn completely and that also change frequently. It should be noted that this distinction is not necessarily directed at other models of cognitive mapping, most of which recognize that human cognitive maps are inherently sketchy, but instead is intended as a general comment on representational strategy. A Gibsonian/knowledge-based/connectionist synthesis recognizes the value of the environment and the information it affords as well as the usefulness of having a model of it. Relying on the environment to a certain extent, for example, in depending on natural gateways and allowing the perceptual system to provide any needed precision, allows for a more economic system than is possible using a knowledge-based approach.

Much of what is unique about PLAN with regard to the cognitive mapping literature is not new with regard to cognitive science. The treatment of landmarks, for example, is based heavily upon prototype theory. The scene-based portion of PLAN has arisen out of work on the human visual system, particularly the location system. The high-level abstractions in PLAN are not only built on the lower level scenes, but draw upon the descriptions of such abstractions in the developmental literature. Nevertheless, although many of the pieces are not new, PLAN represents a new approach to cognitive mapping: an approach that is sensitive both to the ways in which humans process information and to the environment, the source of that information.

## REFERENCES

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A pattern language.* New York: Oxford University Press.

Andersen, R.A., & Zipser, D. (1990). A network model for learned spatial representation in the posterior pareital cortex. In J.L. McGaugh, N.M. Weinberger, & G. Lynch (Eds.), *Brain organization and memory: Cells, systems and circuits.* New York: Oxford University Press.

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

Asada, M., Yasuhito, F., & Tsuji, S. (1988). Representing a global map for a mobile robot with relational local maps from sensory data. *Proceedings of the 9th International Conference on Pattern Recognition,* 520-523.

Bates, E.A., & Elman, J.L. (1992). *Connectionism and the study of change* (Tech. Rep. No. 9202). La Jolla: University of California at San Diego, Center for Research in Language.

Brooks, R.A. (1985). Visual map making for a mobile robot. *Proceedings of the IEEE Conference on Robotics and Automation,* 824-829.

Brooks, R.A. (1991). Challenges for complete creature architectures. In J.A. Meyer & S.W. Wilson (Eds.), *From animals to animats, proceedings of the First International Conference on the Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Bruner, J.S. (Ed.) (1957). On going beyond the information given. In *Contemporary approaches to cognition.* Cambridge, MA: Harvard University Press.

Bushnell, M.C., Goldberg, M.E., & Robinson, D.E. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior pareital cortex related to selective visual attention. *Journal of Neurophysiology, 46* 755-771.

Clark, A. (1989). *Microcognition: Philosophy, cognitive science, and parallel distributed processing.* Cambridge, MA: MIT Press.

Evans, G.W., Smith, C., & Pezdek, K. (1982). Cognitive maps and urban form. *Journal of the American Planning Association,* 232-244.

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

Golledge, R.G. (1987). Environmental cognition. In D. Stokols, & I. Altman, (Eds.), *Handbook of environmental psychology* (Vol. I). New York: Wiley.

Gopal, S., Klatzky, R.L., & Smith, T.R. (1989). NAVIGATOR: A psychologically based model of environmental learning through navigation. *Journal of Environmental Psychology, 9,* 309-331.

Gopal, S., & Smith, T.R. (1990). Human wayfinding in an urban environment: A performance analysis of a computational process model. *Environment and Planning, 22,* 169-191.

Hebb, D.O. (1949). *The organization of behavior.* New York: Wiley.

Kaplan, S. (1973). Cognitive maps in perception and thought. In R.M. Downs & D. Stea (Eds.), *Image and environment.* Chicago: Aldine.

Kaplan, S. (1976). Adaptation, structure and knowledge. In G.T. Moore & R.G. Golledge (Eds.), Environmental knowing: Theories, perspectives, and methods. Stroudsburg, PA: Dowden, Hutchinson & Ross.

Kaplan, S., & Kaplan, R. (1989). *Cognition and environment.* Ann Arbor, MI: Ulrichs.

Kaplan, S., Sonntag, M., & Chown, E. (1991). Tracing recurrent activity in cognitive elements (TRACE): A model of temporal dynamics in a cell assembly. *Connection Science, 3,* 179-206.

Kaplan, S., Weaver, M., & French, R. (1990). Active symbols and internal models: Towards a cognitive connectionism. *AI & Society, 4,* 51-71.

Keil, F.C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior, 23,* 221-236.

Kuipers, B.J. (1978). Modeling spatial knowledge, *Cognitive Science, 2,* 129-153.

Kuipers, B.J., & Byun, Y. (1991). A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Journal of Robotics and Autonomous Systems, 8,* 47-63.

Kuipers, B.J., & Levitt, T.S. (1988). Navigation and mapping in large-scale space. *AI Magazine, 9*(2), 25-43.

Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago: University of Chicago Press.

Leiser, D., & Zilbershatz, A. (1989). The Traveller: A computational model of spatial network learning. *Environment and Behavior, 21,* 435-463.

Lesperance, R.M. (1990). *The location system: Using approximate location and size information for scene segmentation.* Unpublished doctoral dissertation, The University of Michigan, Ann Arbor.

Levenick, J. (1985). *Knowledge representation and intelligent systems: From semantic networks to cognitive maps.* Unpublished doctoral dissertation, The University of Michigan, Ann Arbor.

Levenick, J.R. (1991). NAPS: A connectionist implementation of cognitive maps, *Connection Science, 3,* 107-126.

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

Maes, P. (1993). Behavior-based artificial intelligence. In J.A. Meyer, H.L. Roitblat, & S.W. Wilson (Eds.), *From animals to animats 2, Proceedings of the Second International Conference on Simulation of Adaptive Behavior.* Cambridge, MA: MIT Press.

Mandler, G. (1975). Consciousness: Respectable, useful, and probably necessary. In R.L. Solso (Ed.), *Information processing and cognitive psychology.* Hillsdale, NJ: Erlbaum.

Mataric, M.J. (1990). *A distributed model for mobile robot environment-learning and navigation.* (MIT AI Lab Tech Rep. No. 1228). Cambridge, MA: Massachusetts Institute of Technology.

Miller, H. (1956). The magical number seven, plus or minus two: Some limits on our capacity to process information. *Psychological Review, 63,* 81-97.

Miller, H. (1992). Human wayfinding, environment-behavior relationships, and artificial intelligence. *Journal of Planning Literature, 7,* 139-150.

Newell, A., Shaw, J.C., & Simon, H.A. (1960). Report on a general problem-solving program. In *Proceedings of the International Conference on Information Processing,* Paris: UNESCO.

O'Keefe, J. (1989). Computations that the hippocampus might perform. In L. Nadel, L.A. Cooper, P. Culicover, & R.M. Harnish (Eds.), *Neural connections, mental computation.* Cambridge, MA: MIT Press.

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

O'Neill, M.J. (1990). Computer simulation of the cognitive map: A validation study. In R.I. Selby, K.H. Anthony, J. Choi, & B. Orland (Eds.), *Coming of age: Proceedings of the 21st Annual Conference of the Environmental Design Research Association.* Oklahoma City: EDRA.

O'Neill, M.J. (1991). A biologically based model of spatial cognition and wayfinding. *Journal of Environmental Psychology, 11,* 299-320.

Otto, I., Grandguillaume, P., Boutkhil, L., Burnod, Y., & Guigon, E. (1992). Direct and indirect cooperation between temporal and parietal networks for invariant visual recognition. *Journal of Cognitive Neuroscience, 4,* 35-57.

Pailhouse, J. (1969). Representation de l'espace urbain at cheminements (Representation of urban space and roadways). *Le Travail Humain, 32,* 239-262.

Passini, R. (1984). *Wayfinding in architecture.* New York: Van Nostrand Reinhold Company.

Piaget, J., & Inhelder, B. (1967). *The child's conception of space.* New York: W.W. Norton.

Porteous, J.D. (1971). The quality of the urban environment. *Environment and Behavior, 3,* 155-178.

Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353-363.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Erlbaum.

Rueckl, J., Cave, K., & Kosslyn, S. (1988). Why are "what" and "where" processed by separate cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience, 1,* 171-186.

Sacerdoti, E.D. (1974). Planning in a hierarchy of abstraction spaces. *Artificial Intelligence, 5,* 115-135.

Shemyakin, F.N. (1962). General problems of orientation in space and space representations. In B.G. Anayev, (Ed.), *Psychological Science in the USSR* (Vol. 1). Washington, DC: U.S. Joint Publications Research Service.

Siegel, A.W., & White, S.H. (1975). The development of spatial representation of large-scale environments. In H.W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10). San Diego, CA: Academic.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1-74.

Squire, L.R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys and humans. *Psychological Review, 99,* 195-231.

Ungerleider, I.G., & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, & R.J.W. Mansfield (Eds.), *Analysis of visual behavior.* Cambridge, MA: MIT Press.

Warren, D.H., & Scott, T.E. (1993). Map alignment in traveling multisegment routes. *Environment and Behavior, 25,* 643-666.

Yeap, W.K. (1988). Towards a computational theory of cognitive maps. *Artificial Intelligence, 34,* 297-360.