

Spatial Cognition and Computation

An Interdisciplinary Journal

Editors

Anthony G. Cohn

Stephen C. Hirtle



Lawrence Erlbaum Associates, Publishers
Mahwah, New Jersey London

Spatial Cognition and Computation

An Interdisciplinary Journal

Editors

Stephen C. Hirtle
*School of Information Sciences,
University of Pittsburgh
Pittsburgh, PA 15260, USA*
hirtle@pitt.edu

Anthony G. Cohn
*Division of Artificial Intelligence,
School of Computing,
University of Leeds
Leeds, LS2 9JT, United Kingdom*
agc@comp.leeds.ac.uk

Advisory Editors

Reg Golledge, *University of California at Santa Barbara*
Ben Kuipers, *The University of Texas at Austin*
Lynn Nadel, *University of Arizona*
Barbara Tversky, *Stanford University*

Editorial Board

Gary Allen, *University of South
Carolina*
Michel Denis, *LIMSI-CNRS*
Max Egenhofer, *University of Maine*
Andrew Frank, *Technical University of
Vienna*
Christian Freksa, *University of Bremen*
C.R. Gallistel, *Rutgers University*
Tommy Gärling, *Göteborg University*
Janice Glasgow, *Queen's University*
Christopher Gold, *University of
Glamorgan*
Christopher Habel, *University of
Hamburg*
Ray Jackendoff, *Brandeis University*
Philip Johnson-Laird, *Princeton
University*
Roberta Klatzky, *Carnegie Mellon
University*
Werner Kuhn, *University of Münster*
Barbara Landau, *Johns Hopkins
University*
Stephen Levinson, *MPI Nijmegen*

Gerard Ligozat, *LIMSI-CNRS*
Jack Loomis, *University of California
at Santa Barbara*
David Mark, *University of Buffalo*
Timothy P. McNamara, *Vanderbilt
University*
Daniel R. Montello, *University of
California at Santa Barbara*
Amitabha Mukerjee, *Indian Institute of
Technology*
John O'Keefe, *University College
London*
Herb Pick, *University of Minnesota*
Juval Portugali, *Tel Aviv University*
Zenon Pylyshyn, *Rutgers University*
Hanan Samet, *University of Maryland*
Jeanne Sholl, *Boston College*
Elizabeth S. Spelke, *Massachusetts
Institute of Technology*
David Stea, *Southwest Texas State
University*
David Uttal, *Northwestern University*

Technical Editor

Denis Nkweteyim, *University of Pittsburgh*

Visit SCC's website at <http://www.sis.pitt.edu/~scc>

An Artificial Neural Network That Uses Coarse Allocentric Coding of Direction to Represent Distances Between Locations in a Metric Space

Michael R.W. Dawson, Patricia M. Boechler, and Jill Orsten
University of Alberta

An artificial neural network was trained to rate the distances between all possible pairs taken from a set of thirteen different cities on the map of Alberta. After learning to solve this problem, the internal structure of the network was analyzed. It was found that in spite of the fact that the network was trained to rate distances, the hidden units in the network were not sensitive to distance at all. Instead, it was found that each hidden unit could be located at a particular latitude and longitude on the map of Alberta, and that at this location each hidden unit was associated with a specific heading. Each hidden unit measured the bearing of each city relative to the hidden unit's heading, and when a stimulus was presented to the network, the hidden units combined the two city bearings and generated an activity level based on this combination. The output units of the network generated an accurate distance rating by combining the directional signals that were being computed by the hidden units. These results are discussed in terms of their implications for theories about cognitive maps, in particular the modularization of distance measurements from directional measurements.

Keywords: Cognitive map, scope slip, non-metric representation, multilayer perceptron.

Introduction

Representations that preserve the metric properties of space have been fundamentally important to the study of how humans and animals navigate (Kitchin, 1994). Behavioral studies have demonstrated that animal representations of space do indeed appear to preserve a good deal of its metric nature (for introductions, see Cheng & Spetch, 1998; Gallistel, 1990). Since the discovery of place cells in the hippocampus (O'Keefe & Nadel, 1978), many researchers have been concerned with identifying the biological substrates that encode metric space (e.g., Redish, 1999).

The Hippocampus as a Cognitive Map

One key proposal about these biological substrates is that they instantiate a cognitive map of the type proposed by Tolman (1948). For Tolman, a cognitive map indicated "routes and paths and environmental relationships" (p. 192). One of its key characteristics was that it permitted an animal to make novel shortcuts between locations in space (Bennett, 1996). However, while Tolman made strong arguments for the existence of cognitive maps, he did not discuss their primitive properties to any great extent.

O'Keefe and Nadel (1978) developed Tolman's (1948) hypothesis in much more detail, and proposed that the hippocampus was the neural substrate that mediated the cognitive map. They argued that the cognitive map was an innate Euclidean spatial framework, and was thus governed by the basic properties of metric space. This framework was proposed as an absolute space that existed independently of any objects that could be located within it. Objects were located in this space by taking as input a number of different egocentric configurations that the animal experienced as it explored the environment (taxon spatial configurations) and placing them into a cohesive global representation. This global representation was an allocentric map in the sense that it could provide information about the distance and direction between any pair of places that were encoded, independently of an animal's own location or orientation in the space.

Neurons in the hippocampus were seen as being ideally suited for implementing a cognitive map with these sorts of properties. O'Keefe and Nadel (1978) described the properties of hippocampal place cells that respond only when a rat's head is in a particular location in the environment. These place cells can be driven by visual information (e.g., by the presence of objects or landmarks in the environment), and appear to be sensitive to some of the metric attributes of space (e.g., O'Keefe and Burgess, 1996).

O'Keefe and Nadel (1978) also discussed the properties of hippocampal place cells that are associated with the behavior of an animal independent of its location in the environment. These cells exhibit an overall increase in firing rate when an animal is engaged in a variety of behaviors that change its position in the environment (i.e., that displace the animal's position). O'Keefe and Nadel point out that the behavior of these cells has been linked to the theta rhythm observed in EEG studies of the hippocampus. Theta rhythm is associated with

the same set of behaviors associated with the firing of displace cells, and the frequency of theta rhythm may reflect the distance traveled by an animal as it moves. In other words, displace cells might be capable of representing distance traveled in a space that is being explored.

O'Keefe and Nadel (1978) proposed that the hippocampus could construct a cognitive map via the coordinated activities of place cells and displace cells. When the animal initiates some exploratory behavior from its location at place A, the place cell representation of this location is active. As it moves from place A to some goal place B, theta activity from the displace cells is generated. With the animal's movement, the peak of this theta activity will shift its location in the CA3 field of the hippocampus, causing associative processes to engage different place cells. When the animal reaches place B, a path from A to B in the map will be represented as an ordered set of place cells that were associated with one another during the exploration. Because the frequency of theta activity changes with the speed of the animal's movement, the same path of place cells would be activated whether the animal was moving quickly or slowly, suggesting that the path would represent the distance between A and B relatively independently of the animal's speed. In short, combining place cell and displace cell activity permits the neural representation of locations in space, as well as the relative distances between these locations, resulting in a cohesive spatial map.

More recently, some researchers have elaborated and modified O'Keefe and Nadel's (1978) proposal. For example, it has been argued that hippocampal place cells make up only a part of the cognitive map, and that the neural representation of metric space requires the coordination of a number of different functional subsystems that are mediated by a number of different areas in the brain (e.g., McNaughton et al., 1996, Redish & Touretzky, 1999; Touretzky, Wan, and Redish, 1994, Worden, 1992). In one such proposal, Redish (1999) hypothesized that animal navigation depends upon five different representational systems. These include a local view that represents an animal's relationship to environmental landmarks, a path integration system that supports vector arithmetic, a head direction system that encodes orientation in space, a place code that coordinates local views with path integration coordinates, and a goal memory system that associates spatial inputs with motivational states and is used to plan trajectories. Each of these different systems is characterized by different representations that can include both spatial and nonspatial information. These different functional components mediated by the hippocampus as well as other areas of the brain. Nevertheless, the key elements of O'Keefe and Nadel's proposal persist in this theory and others like it. "Within all these accounts, cognitive maps involve a systematic, cohesive, allocentric coordinate framework into which environmental features are encoded" (Eichenbaum, Dudchenko, Wood, Shapiro, & Tanila, 1999).

Arguments Against the Cognitive Map Hypothesis

Importantly, a number of researchers have argued against the cognitive map hypothesis, and against the proposal that the hippocampus provides a metric,

systematic, and cohesive cognitive map. First, place cells are not organized topographically; the arrangement of place cells in the hippocampus is not isomorphic to the arrangements of locations in an external space (Burgess, Recce & O'Keefe, 1995; Eichenbaum et al., 1999; McNaughton et al., 1996). Second, place cell receptive fields (place fields) are at best *locally* metric (Touretzky, Wan, & Redish, 1994), and as a result a good deal of spatial information (e.g., information about bearing) cannot be derived from place cell activity. Third, the responses of different place cells do not appear to be related to one another in a coherent and holistic spatial framework (Eichenbaum et al., 1999). Fourth, hippocampal cells appear to be sensitive to a wide variety of nonspatial variables (Best, White, & Minai, 2001; Eichenbaum, 2002, chap. 6). "Multiple lines of evidence are inconsistent with the idea that hippocampal cells are dedicated to spatial coding and instead indicate that their scope of representation extends to the full range of regularities present in the experience" (Eichenbaum et al., 1999).

Evidence of this sort has led some to abandon the view that the hippocampus provides a spatial cognitive map. For instance, Bennett (1996) suggests that "the cognitive map is no longer a useful hypothesis for elucidating the spatial behavior of animals, and that the use of the term should be avoided." Eichenbaum (2002; Eichenbaum et al., 1999) takes a more constructive approach to replacing the cognitive map hypothesis by proposing that the hippocampus does not represent space per se, but rather stores episodic events in a memory space. In the memory space theory, such cells "are simply an example of the nodal codings that can identify past episodes that share a common event—in this case, a 'place' experienced in the past" (Eichenbaum et al., 1999).

Cognitive Maps and the Scope Slip

Nonspatial theories like those of Eichenbaum (2002) are important, plausible and interesting. However, abandoning the cognitive map hypothesis because hippocampal properties are not isomorphic to those of metric space could also mean falling victim to what Pylyshyn (1981) has called the *scope slip*. Pylyshyn introduced the scope slip as part of the imagery debate. According to Pylyshyn, researchers often mistakenly assume that properties of represented content are also properties of the underlying representation. That is, when the scope slip occurs in the study of mental imagery, researchers note that the content of mental images has spatial properties, but go on to mistakenly assume that the representational medium that forms mental images has these spatial properties as well.

With respect to theories about the cognitive map, the scope slip occurs when researchers assume that a represented space has particular properties, and then mistakenly assume that the substrate that represents this space must also have these properties. To parallel Pylyshyn (1981), let us first note that many theories that focus on space and navigation are concerned with providing accounts of how the hippocampus provides a "representation of space X with properties P." For O'Keefe and Nadel (1978), the properties would be those of

an absolute Euclidean space. Furthermore, these spatial properties are assumed to be properties of the represented space, and not properties of the representing medium. "The cognitive map is not a picture or image which 'looks like' what it represents; rather, it is an information structure from which map-like images can be reconstructed and from which behavior dependent upon place information can be generated" (O'Keefe & Nadel, 1978). In other words, O'Keefe and Nadel assume that the hippocampus is a *representation of (space X with properties P)*, where the parentheses in this statement indicate that "space X" is the component that possesses *properties P*.

However, as was noted earlier, critics of the cognitive map hypothesis cite evidence that the hippocampus does not exhibit the same properties as the space being represented (e.g., topographic organization). "There is no strong evidence that these place representations are elements of a cohesive map of space" (Eichenbaum, 2002). However, this argument makes the scope slip because it relies on the claim that the hippocampus is a "(representation of space X) with properties P." Note the change in location of the parentheses in this expression, indicating that it is the *representation*, and not just the space, that has *properties P*. This expression fails to recognize O'Keefe and Nadel's (1978) assumption that while the hippocampus represents a space with certain properties, the hippocampus does not itself have to exhibit these same properties.

An Example Non-Metric Representation of Metric Space

The possibility of making the scope slip is important to keep in mind, because it can be shown that nonspatial representations are capable of representing a metric space. For instance, Dawson, Boechler, and Valsangkar-Smyth (2000) trained an artificial neural network (ANN) to rate the distances between all possible pairs of thirteen different city locations taken from the map of Alberta. The ratings were based on measured distances between cities on a road map of the province, and the entire set of ratings could be used to derive an accurate two-dimensional graph of the locations of the thirteen cities. Thus, when the network learned to correctly generate this set of ratings, it could be said to have internalized a metric map, in O'Keefe and Nadel's (1978) sense that network was a system from which map-like information could be reconstructed.

However, when Dawson et al. (2000) examined the internal structure of this network, they found that its internal components (i.e., the *hidden units* in the network) were not very map-like. Each hidden unit could be assigned a location on the map of Alberta such that there was a high degree of correlation between connection weights and the distances between the unit and each city. Nevertheless, this distance information was highly distorted, so that the responses of each hidden unit were at best locally metric. However, when several different patterns of such distorted representations were combined, an accurate spatial judgment (i.e., a distance rating) could be generated. Dawson et al. called this representational scheme coarse allocentric coding. Concerning the scope slip, coarse allocentric coding is a (*representation of space X*) with *properties*

P," but is not a *representation of (space X with properties P)*, where *properties P* are understood to be the characteristics of a metric space.

The Synthetic Approach to Cognitive Maps

The previous example shows that avoiding the scope slip is important in the study of cognitive maps, because a metric space can be represented by a medium that does not exhibit metric properties itself. Given this situation, it is perhaps not surprising that some researchers use a much broader definition of a cognitive map to guide their research on navigation. For example, Gallistel (1990, p. 103) defines a cognitive map as "a record in the central nervous system of macroscopic geometric relations among surfaces in the environment used to plan movements through the environment." Such a definition is neutral concerning the specific neural properties of a cognitive map, and about the spatial properties that it encodes. However, it can be used to motivate two general types of research inquiries.

The first type of question concerns investigating cognitive maps at what could be called the computational level (e.g., Dawson, 1998). At the computational level, researchers are interested in determining what information processing problems are solved by cognitive maps. This would include identifying the general properties of the space represented by the map (e.g., is it Euclidean? Is it metric?), the specific kinds of spatial attributes encoded in the map (e.g., is distance encoded? Is direction encoded?), and the general nature in which these spatial attributes are initialized and combined (e.g., How is direction set in the map? How are distance and direction combined to mediate navigation?). In general, computational approaches involve formal accounts of information processing, and as a result Bennett (1996) endorses Gallistel's (1990) more liberal definition of cognitive maps. Bennett points out that computational analyses can take advantage of mathematicians' formal treatment of space, and can test the properties of an animal's cognitive map against precise predictions taken from geometry.

Gallistel is primarily concerned with answering computational questions about cognitive maps. For instance, Gallistel (1990) noted, "a central question is what type of geometric relations a map encodes." He has reviewed a great deal of literature on animal navigation (Gallistel, 1990, chap 4-6; 1998) that supports the claim that even insects are capable of determining angles and distances, of performing dead reckoning, of holding a course, and of recognizing a terrain from different perspectives and distances. Gallistel goes on to argue that these abilities are mediated by an integrated cognitive map, and that bees, wasps, and other insects use this map to solve complicated geometric problems in order to navigate through their environment.

A second kind of question that is motivated by Gallistel's (1990) broader definition of cognitive map has been called the algorithmic question (e.g., Dawson, 1998). If, at the computational level, one can make the case that a particular information processing problem is being solved, and that the solution to this problem involves manipulating particular variables, then the next step is

to consider the specific manner in which these variables are represented, and the specific operations that are used to process them. Later, once the algorithmic questions have been answered, one can turn to answering the implementational question by discovering how particular representations and operations are instantiated by the brain.

Gallistel (1998) recognizes the importance of exploring algorithmic and implementational questions. "If we are persuaded by the behavioral evidence that the nervous system really does compute, then that establishes an agenda for behavioral neuroscience, namely, to discover the processes that enable it to do so. What processes enable it to store the value of a variable?" However, Gallistel has little to say about the cognitive map at the algorithmic or the implementational level. This is because he believes that the regularities that he has proposed at the computational level pose problems for investigators concerned with algorithmic or implementational issues. He argues that this is particularly true for algorithmic accounts that are based upon artificial neural networks, because "we do not know with any certainty how the nervous system implements even the operations that are assumed in neural net models, let alone the fuller set of operations that are taken for granted in any computational/symbolic model of the processes that mediate behaviors" (Gallistel, 1998).

The obvious response to this position is to take an artificial neural network and use it to study cognitive maps by adopting the methods of synthetic psychology (Braitenberg, 1984; Dawson, 2004; Pfeiffer & Scheier, 1999). When artificial neural networks are used in this approach, one first trains the network to behave in a particular way. For instance, one can train it to generate a set of responses that support the claim that the network has internalized a particular set of spatial relations. After this training has been accomplished, one examines the network's internal structure in an attempt to discover how these spatial relations are represented in the network. The hope is that the analysis of the network's internal structure will reveal novel types of representations that can be used to inform theories of cognitive maps.

Dawson et al.'s (2000) network, described above, provides one illustration of this synthetic approach. Because the network was able to make accurate distance judgments between pairs of cities, it was clear that the network had internalized a map that represented some of the spatial relations evident on a road map of Alberta. However, its internal structure was far from map-like. The coarse allocentric coding that was discovered inside the network indicated one manner in which distance could be represented in a network (i.e., encoded as connection weights feeding into hidden units), and showed how judgments of the distance between a pair of cities could be computed by sensing the distances between two cities and a number of different locations on the map of Alberta that were not themselves occupied by cities. Dawson et al. (2000) also found that coarse allocentric coding was flexible enough to represent a nonmetric space in which the minimality principle was violated.

The purpose of the current paper is to describe another network that illustrates this synthetic approach. The network was trained on the same Alberta map problem that was described by Dawson et al. (2000), but used a different input encoding. To preview the main findings that are reported below, the internal representations of this new network also exhibited coarse allocentric coding. However, in spite of the fact that the network was trained to make distance judgments, the network's hidden units did not represent distance. Instead, each hidden unit represented angles between its point of view and the various cities on the map, and these measurements of direction were used to compute distance.

Method

Problem Definition

In the simulation described below, we trained a network to make judgments about the "crows flight" distances between pairs of places on a map. We chose thirteen different locations in the province of Alberta: Banff, Calgary, Camrose, Drumheller, Edmonton, Fort McMurray, Grande Prairie, Jasper, Lethbridge, Lloydminster, Medicine Hat, Red Deer, and Slave Lake. We took all possible pairs from this set to create a set of 169 different stimuli, each of which could be described as the question "On a scale from 0 to 10, how far is City 1 from City 2?" Because all possible pairs of place names were used, thirteen of the stimuli involved rating the distance from one city to itself (e.g., the distance between Banff and Banff). As well, for different place names a rating would be obtained for both orders of places (e.g., the distance between Banff and Calgary would be rated, as would be the distance between Calgary and Banff).

The desired ratings for each stimulus were created as follows. First, from a map of Alberta we determined the shortest distance in kilometers between each pair of locations (see Dawson et al., 2000, Table 1a). Second, these distances were converted into ratings. If a stimulus involved rating the distance from one place to itself, the rating was assigned a value of 0. Otherwise, if the distance was less than 100 kilometers, then it was assigned a value of 1; if the distance was between 100 and 199 kilometers, then it was assigned a value of 2; if the distance was between 200 and 299 kilometers, then it was assigned a value of 3; and so on up to a maximum value of 10 which was assigned to distances of 900 kilometers or more. The complete set of ratings that were used is provided in Dawson et al. (2000, Table 1b).

Because this ratings matrix was based on converting continuous physical distances into discrete ratings, it is important to confirm that the ratings data does not dramatically distort the original distances. Dawson et al. applied multidimensional scaling to the ratings data, forcing a solution in which the configuration of points delivered by MDS was restricted to two dimensions. The solution accounted for 99.4% of the variance in the ratings matrix, and produced a configuration of points that agreed quite nicely with the locations of the thirteen cities on a road map of Alberta.

Network Architecture

An ANN was trained to make the distance ratings that were described above. The input units were used to represent a pair of cities, and the output units were used to encode discrete ratings of the distances between the input cities. The network had 10 output units, 7 hidden units, and 26 input units, and is illustrated

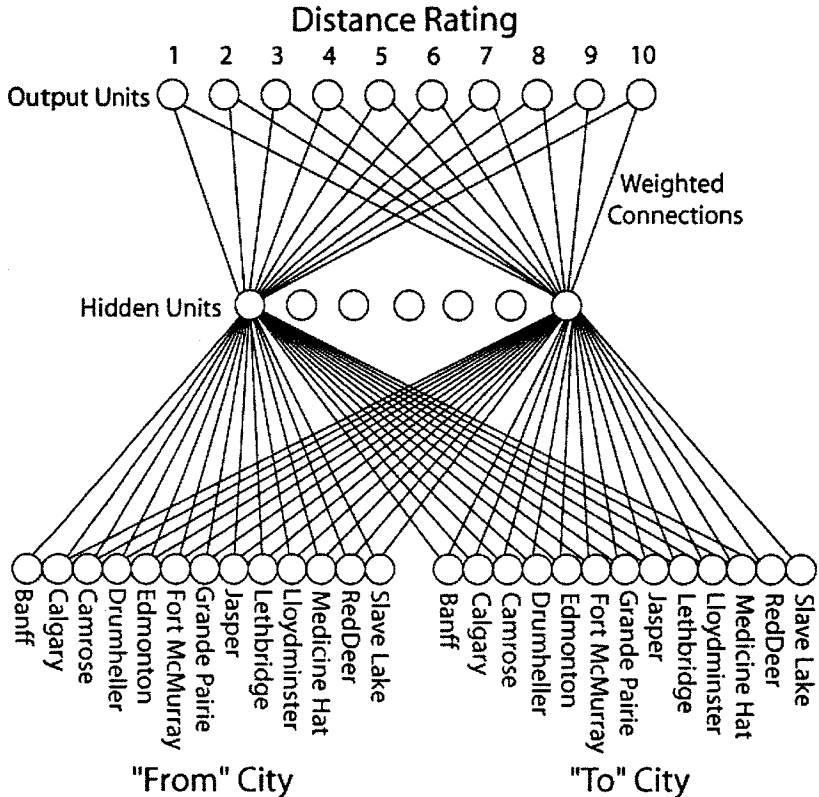


Figure 1. The multilayer perceptron used in this study. Circles represent processing units, and lines represent modifiable connections between processors. In the bottom layer of input units, each unit represents a city name. When a stimulus is presented, only two of these units are turned on—one *from city* name, and one *to city* name. The input signals are then processed by 7 hidden units in the next layer, which then send signals to the 10 output units. After training, the network will only turn one of its output units on, representing its rating of the distance between the *from city* and the *to city*. For the sake of simplicity, the figure only depicts the connections involving two of the hidden units. In the actual network, every input unit was connected to every hidden unit, and every hidden unit was connected to every output unit.

in Figure 1. All of the output units and all of the hidden units were value units that used the Gaussian activation function that was described by Dawson and Schopflocher (1992).

ANN architectures can differ from one another along a variety of dimensions (for an introduction, see Dawson, 2004). One important distinction between different classes of ANNs involves the activation function that converts a unit's total input into an internal level of activity (Duch & Jankowski, 1999). Typically, ANNs are composed of units whose activation function is a sigmoid-shaped curve that is defined by the logistic equation; Ballard (1986) calls such units *integration devices*. However, not all networks are composed of integration devices. Some networks use processors that are tuned to activate to a small range of net inputs, and generate weak responses to net inputs that are either too small or too large to fall in this range. Ballard calls such units *value units*.

All of the output units and all of the hidden units in the network that was trained were value units, and used the Gaussian activation function that was described by Dawson and Schopflocher (1992): $G(net_i) = \exp(-\pi(net_i - \mu_i)^2)$. In this equation, $G(net_i)$ is the activation being calculated for output unit i , net_i is the net input for that output unit, and μ_i is the mean of the Gaussian. When the net input to the Gaussian equation is equal to the mean (i.e., equal to μ_i), the activity that is generated is equal to 1.0. As a result, μ_i can be thought of as being similar to the bias of the logistic or the threshold of the step function.

While the value units used by Dawson and Schopflocher (1992) employ the Gaussian activation function, they are quite distinct from other units that employ Gaussian equations. In particular, value units are not radial basis function (RBF) units. This is because value units and traditional integration devices (e.g., Rumelhart, Hinton, and Williams, 1986) use the same net input function (an inner product between a vector of weights and a vector of unit activities). Other Gaussian-based units—in particular, RBF units—use a nonstandard net input function, which involves computing the distance between the position of an RBF unit and an input pattern when both are viewed as occupying different positions in a pattern space. A more detailed discussion of the difference between value units and RBF units is provided in Dawson and Schopflocher (1992).

A network of value units was used in this study for two reasons. First, a number of different studies have demonstrated that networks of value units permit their internal structure to be interpreted in great detail (Berkeley, Dawson, Medler, Schopflocher, & Hornsby, 1995; Dawson, 1998, 2004; Dawson et al., 2000; Dawson, Medler, & Berkeley, 1997; Dawson, Medler, McCaughan, Willson, & Carbonaro, 2000; Dawson & Piercey, 2001; Dawson & Zimmerman, 2003; Leighton & Dawson, 2001). Second, place cells in the hippocampus behave as though they are value units—they are tuned to respond to particular ranges of spatial measurements in the environment, and their behavior suggests that their distance sensitivity is modulated by a Gaussian response function (e.g., O'Keefe & Burgess, 1996). This suggests that the value unit architecture is particularly appropriate for studying how networks of parallel processors

represent spatial information, particularly if one goal of the simulation research is to explore representations that might be found in hippocampal circuitry.

Input Unit Representation

The 26 input units illustrated in Figure 1 were used to define a local code for pairs of cities to be compared. Each input unit represented a place name; the first input unit represented Banff, the second input unit represented Calgary, and so on alphabetically.

Every stimulus presented to the network was a request to rate the distance "from City A to City B." Thus two different city names had to be represented in an input pattern: the name of the *from city* (i.e., the first name in a pair of names), and the name of the *to city* (i.e., the second name in a pair of names). The first thirteen input units represented the name of the *from city*, while the remaining thirteen input units represented the name of the *to city*. As a result, each pattern that was presented to the network involved turning two input units on, one that belonged to the first set of thirteen input units, and one that belonged to the second set of thirteen input units. For example, to ask the network to judge the distance from Banff to Calgary, the first and the fifteenth input units would be turned on, and all other input units would be turned off. Similarly, to ask the network to judge the distance from Calgary to Banff, the second and the fourteenth input units would be turned on, and all other input units would be turned off.

This representational scheme was chosen because it contains absolutely no information about the location of the different places on a map of Alberta. In other words, the input units themselves did not provide any metric information that the network could use to perform the ratings task. All possible pairs of names from the set of thirteen cities were used as training patterns. As a result, there were 169 different input patterns that were used to train the network.

Output Unit Representation

Ten output units were used to represent the network's rating of the distance between the two place names presented as input. The output units were also value units. To represent a rating of 0, the network was trained to turn all of its output units off. To represent any other rating, the network was trained to turn on one, and only one, of its output units. Each of these output units represented one of the ratings from 1 to 10. For example, if the network turned output unit 5 on, this indicated that it was making a distance rating of 5.

Number of Hidden Units

Seven hidden units were included in this network to solve the problem (see Figure 1). Each of these units was a value unit. We selected this number of hidden units because pilot tests had shown that this was the smallest number of hidden units that could be used by the network to discover a mapping from input to output. When fewer than seven hidden units were used, the network was never able to completely learn the task. Previous research has suggested that forcing a network to learn a task with the minimum number of hidden units

produces a network that is much easier to interpret, in comparison to a network that has more hidden units than are required to solve a problem (Berkeley et al., 1995).

Network Training

Typical multilayer ANNs are trained to accomplish a desired pattern classification task by using the generalized delta rule (Rumelhart et al., 1986). The network in the current paper was trained using the variation of the generalized delta rule that has been developed for networks of value units (Dawson & Schopflocher, 1992). This variation of the learning rule performs a gradient descent in an error space, as does the traditional generalized delta rule. However, it uses an elaborated definition of output unit error to prevent networks of value units from stalling at local minima. It can be shown that when the Dawson and Schopflocher learning rule optimizes the elaborated definition of error, it also optimizes the traditional definition of error. In other words, the two learning rules arrive at the same minimal value of error at the end of training. The similarities and differences between these two types of rules are discussed in detail in chapters 10 and 11 of Dawson (2004).

Prior to training, all of the connection weights were randomly assigned values ranging from -0.10 to +0.10. The biases of processing units (i.e., the means of the Gaussian activation functions, which are analogous to thresholds) were also randomly assigned values ranging from -0.10 to +0.10. The network was trained with a learning rate of 0.005 and zero momentum. During each sweep of training, each of the 169 stimuli was presented to the network, with the order of presentation of the set of 169 patterns randomized prior to beginning the sweep. The learning rule was used to update connection weights in the network after each stimulus presentation.

Training proceeded until the network generated a *hit* for every output unit on every pattern. A hit was operationalized as an activation of 0.90 or higher when the desired activation was 1.00, and as an activation of 0.10 or lower when the desired activation was 0.00. The network converged on a solution to the problem—generating a correct response for each of the 169 patterns—after 5078 sweeps of training.

Results

When artificial neural networks are used as a medium for synthetic psychology, what is of primary interest is the interpretation of their internal structure (e.g., Dawson, 2004). The hope is that the network has developed a novel representation for mediating stimulus/response regularities. However, networks are generally difficult to interpret. The sections that follow describe a number of different strategies that were used to discover the internal representations of the Alberta map network.

Projecting Cities Onto One-Dimensional Weight Spaces

One important source of information that can be used to interpret the internal structure of an artificial neural network is the set of connection weights that feed into each hidden unit. We began to study these weights by making a graph for each hidden unit that represented the location of each city in a two-dimensional space. This type of graph could be easily constructed because, for each hidden unit, every city was associated with two connection weights: its *from weight* (i.e., its weight from one of the first thirteen input units) and its *to weight* (i.e., its weight from one of the second thirteen input units). We simply used these two weights as the coordinates of a city in the graph. Figure 2 provides an example of two of these graphs.

As can be seen from Figure 2A, when the cities are plotted in terms of their pair of connection weights for a hidden unit, the resulting graph is strikingly linear. This was true for the graphs for all seven hidden units. The only exception was Hidden Unit 3. For this hidden unit, 12 of the cities were arranged linearly, but the remaining city (Banff) fell slightly off the line, as is illustrated in Figure 2B.

After noting the linear structure of each of these graphs, our next step was to determine the equation of the line for each graph. For each hidden unit, multiple regression was used to predict a city's *to weight* from its *from weight*. The results of these regressions are presented in Table 1. From this table, two observations can be made. First, the regression lines provide near perfect fit for all of the data. Second, the slopes of each regression line are near unity. The second point makes sense because of the symmetric nature of the training set. For every pair of cities in the training set, the rating of the distance from City A to City B was identical to the rating of the distance from City B to City A. Because of this, it would be expected that the *from weight* for a city should be nearly identical to that city's *to weight*. The regression equations simply confirm this expectation.

The regression equations are useful because they permit us to convert the two-dimensional space of the type depicted in Figure 2 into a one-dimensional space that we will call a *weight space*. This was accomplished as follows: For each graph of the type illustrated in Figure 2, we created a unit length vector from the origin of the graph that had the same slope as the regression equation that was computed for the same graph. For each city, we then calculated the orthogonal projection of the city from its location in the two-dimensional graph onto the vector. The resulting weight spaces for each hidden unit are provided in Table 2.

The Weight Spaces Do Not Encode Distance

The previous section indicated that the weights that fed into each hidden unit were highly systematic. What is the source of this regularity? What do the weights represent? Is there a relationship between the map of Alberta and these connection weights?

Dawson et al., (2000) were able to relate map distances to connection weights. They did this by assuming that each hidden unit occupied a place on the map.

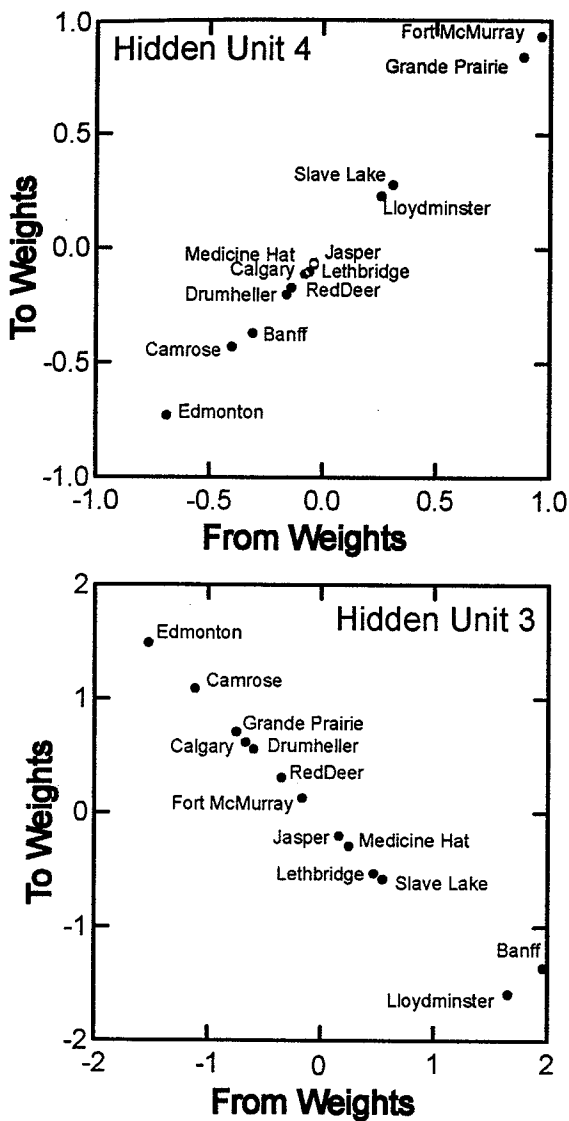


Figure 2. Graphs of cities plotted in terms of their weights feeding into hidden units. (A) The graph for Hidden Unit 4. (B) The graph for Hidden Unit 3. Note in this latter graph the one departure from linearity at the bottom right of the plot, where Banff does not fall in the same line as do the other cities.

Table 1

Results of regressions that predicted a city's weight from its weight for each of the seven hidden units

Hidden Unit	Slope	Interce pt	R2
H1	0.994	0.032	1.000
H2	0.999	-0.007	1.000
H3	-0.885	0.017	0.978
H4	1.004	-0.035	1.000
H5	1.004	0.039	1.000
H6	-0.994	0.018	1.000
H7	-1.000	-0.014	1.000

Table 2

Weight spaces for each hidden unit computed by orthographically projecting cities from a 2D representation based on connection weights onto a vector with the same slope as the regression line predicting to weights from from weights

City	Hidden Unit						
	H1	H2	H3	H4	H5	H6	H7
Banff	0.665	-0.228	-2.368	-0.480	0.483	0.892	-0.524
Calgary	-0.632	-2.333	0.913	-0.135	-0.318	-0.723	0.331
Camrose	0.010	-1.339	1.564	-0.583	-1.252	0.568	-0.971
Drumheller	-0.931	-0.299	0.819	-0.251	0.028	-0.752	-0.822
Edmonton	-0.239	0.341	2.130	-1.005	0.659	0.597	0.987
Fort McMurray	-0.525	0.757	0.216	1.338	0.752	-0.590	-0.428
Grande Prairie	-0.177	0.039	1.027	1.213	-0.108	0.110	-0.667
Jasper	0.796	1.862	-0.257	-0.075	-0.783	0.101	-1.193
Lethbridge	0.612	1.366	-0.707	-0.222	1.215	0.465	0.920
Lloydminster	0.414	0.801	-2.286	0.348	0.818	-0.614	0.332
Medicine Hat	-0.516	1.404	-0.384	-0.071	1.414	0.205	1.764
RedDeer	0.122	-1.924	0.465	-0.113	-2.198	-0.183	-0.351
Slave Lake	0.864	-0.812	-0.802	0.420	-1.185	-0.222	-0.056

They used the Solver tool in Microsoft Excel to change the coordinates of each hidden unit until a maximum value was reached for the correlation between the distances from each city to the hidden unit and the weights of the hidden unit. When they did this, they were able to find a location for each hidden unit that generated a strong correlation; the correlations between distance and weight ranged from 0.48 to 0.88 for the six hidden units in that network. When the map of Alberta was transformed to reflect the fact that the hidden units of that network distorted distance by applying a Gaussian activation function, the correlations became even higher.

Under the assumption that the current network had internalized spatial relations using a similar representational scheme, we repeated this analysis on it. To our surprise, the correlations that we obtained were very poor, ranging from 0.03 to 0.34. This suggests that in spite of the fact that the network was being

trained to make distance judgments, it did not use internal representations of distance to accomplish this task!

The Weight Spaces Encode Direction

If the weights feeding into the hidden units do not represent distance, then what do they represent, and how can this variable be used to make distance judgments? In general, successful navigation depends upon two different kinds of measurements: distance and direction (e.g., Gallistel, 1990, chap. 3). We hypothesized that if the hidden units did not appear to represent the relative distances between cities, then perhaps they were representing some measure of relative direction.

How could measures of direction be used to derive judgments of relative distances on a map? Imagine being on a flat prairie with a small number of distinct landmarks (e.g., trees) in view, and having the task of determining the distances between all of these landmarks. However, the only tool available was a sextant, so that the only measurement that could be taken was the angular displacement between pairs of landmarks. One could make a rough estimate of the distances between landmarks by standing at one location on the prairie and taking a sextant reading between every possible pair of landmarks. The reasoning would be that if a sextant reading was high, then the two landmarks were far apart, and that if it was near zero, then the two landmarks were near one another.

The problem with this approach is that it depends upon the point of view. When standing on the prairie, and looking at the landmarks from a particular direction, two of the landmarks might appear to be close together, when in reality they are far apart. For instance, two landmarks could be distant from one another in the same direction from which they are being viewed, which would mean that they would provide a very small sextant reading in spite of their actual long distance apart. In short, this means that one set of sextant readings would provide inaccurate estimates of the distances between pairs of landmarks.

To correct this problem, one could move to different locations on the prairie, taking a different set of sextant readings from different points of view. Individually, each set of readings would be inaccurate. However, if the distance judgments based on each set of readings were combined in some way, then much higher accuracy would be obtained. This is because the errors in one set of distance estimates, due to the point of view from which they were derived, could be reduced by other estimates derived from other points of view. This is the essence of coarse coding, a process by which highly accurate measurements can result by combining the outputs of a number of less accurate measurements from different perspectives.

Our specific hypothesis, then, was that each hidden unit was measuring angular displacement, and was encoding this displacement in its connection weights. To test this hypothesis, we first assumed that each hidden unit could be assigned a location (i.e., a latitude and a longitude) on the map of Alberta. We then assumed that in addition to this location, each hidden unit had a heading:

that is, it was looking at the landmarks on the map (i.e., the cities) from a particular direction. We operationalized this by assuming that each hidden unit served as the origin for a unit-length vector that was oriented in a fixed direction. Finally, we operationalized the notion of the hidden unit measuring the angular displacement of a city as the cosine of the angle between the hidden unit's heading and the vector from the hidden unit to the city (the bearing of the city).

We tested this hypothesis by using the Solver tool in Microsoft Excel to optimize the correlation between the projections of the cities on to a hidden unit's weight space (Table 2) and the cosines of the angles between the hidden unit's heading vector and the bearing vectors to each of the thirteen cities. We did this as follows. First, we took a hidden unit (with its heading vector pointing north) and positioned it at the centroid of the map of Alberta, where the centroid was the average latitude and longitude of our thirteen cities. Second, we used solver to change the location of the hidden unit to a position that optimized the correlation between the weight space projections and the cosines, with the heading vector pointing in a fixed direction. Third, we held the hidden unit fixed at this new location, and had Solver rotate the heading vector to a new direction that further optimized the correlation between weight space and cosines. We iterated this process (i.e., moving the coordinates of the hidden unit with the heading vector fixed, and then rotating the heading vector with the coordinates of the hidden unit fixed) until no further optimization of the correlation was observed.

The main result of this analysis is reported in Table 3. It shows the correlations between the weight space projections and cosines for each hidden unit after the location and bearing of the hidden unit has been optimized. It is clear from this table that there is a very strong relationship between this particular measure of angular displacement on the map of Alberta and the connection weights that feed into the hidden units. In other words, a hidden

Table 3

Location of each hidden unit on the map of Alberta, location of the endpoint of each hidden unit's heading vector, and the correlation between each unit's connection weights and the cosines of the angles between city bearing vectors and the heading vector

Hidden Unit	Location of Unit		Endpoint of Heading Vector		Correlation
	Latitude	Longitude	Latitude	Longitude	
H1	51.100	114.018	50.717	114.941	0.61
H2	49.173	115.076	50.128	114.779	-0.73
H3	53.283	110.001	53.359	110.999	0.56
H4	56.940	117.837	57.824	118.305	0.87
H5	51.598	112.407	51.045	111.573	0.72
H6	51.100	114.017	51.773	113.277	-0.59
H7	50.894	112.025	51.605	112.729	-0.75

unit's weights are analogous to cosines between the heading of the hidden unit and the bearings of the various cities (e.g., the cosines of angles A and B in Figure 4 which is presented later). The poorest correlation was the 0.56 observed for Hidden Unit 3. However, this was also the one hidden unit that had the one departure from linearity (i.e., Banff as an outlier, see Figure 2B) when *from weights* were plotted as a function of *to weights*. If Banff is omitted from the calculation of the correlation for this unit, and the correlation is based on the cosines to the remaining 12 cities, then the correlation for this unit rises to 0.75. In short, the evidence suggests that each hidden unit can be viewed as a device that is measuring angular displacement and encoding the result in its connection weights.

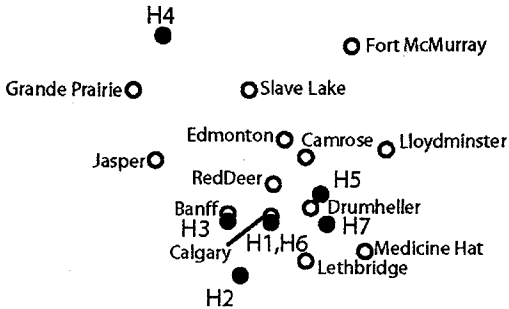
The statistical significance of these correlations can be determined by computing the F -statistic for each correlation using r^2 as the numerator and $(1 - r^2)$ as the denominator, and dividing the numerator and denominator by appropriate degrees of freedom (in this case, by 1 and 11 respectively). When this is done, all of the correlations are statistically significant at $p < 0.05$ or better; the F -value for the smallest correlation is 5.03 and the F -value for the largest correlation is 34.25.

It is important to consider the size of the correlations in relation to the size of the correlation that would result if random data were being fit, because the correlations presented in Table 3 are determined by searching through a parameter space. We replicated the statistical analyses reported above, but this time replaced the actual connection weights of the network with random numbers selected from a range between -1 and 1. When these random weights were fit to map using the procedure described above, the average optimal correlation that was obtained was 0.34. A correlation of this magnitude leads to an F -statistic of 1.44, which is not statistically significant.

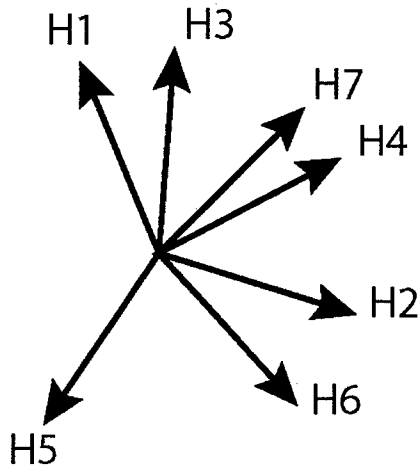
We argued earlier that if measures of angular displacement were used to derive judgments of the relative distance between pairs of landmarks, then a variety of different points of view would be required to make these judgments accurate. Figure 3 illustrates that this is true for the hidden units of this network. Figure 3A uses the coordinate information from Table 3 to plot the position of each hidden unit on a map relative to the actual locations of the thirteen cities. It can be seen from this figure that hidden units can be found in a variety of locations on the map, with more of the hidden units being concentrated in the area of the map occupied by the majority of cities. Figure 3B illustrates the 7 heading vectors of the hidden units, after they have all been centered to a common origin. This figure illustrates that in addition to occupying a variety of locations, the hidden units also have a variety of points of view (headings) on the cities.

Converting Bearing Measurements to Net Input

The results presented in the previous section show that the individual connection weights that feed into a hidden unit are related to the angle between the unit's



(A)



(B)

Figure 3. (A) Locations of hidden units relative to cities on the map of Alberta. Hidden units are indicated with solid dots. (B) Headings of each hidden unit, centered to a common origin.

heading and a city's bearing. We can now examine how each hidden unit uses this information to determine its response to an input pattern.

When any stimulus pattern is presented to the network, signals are sent through only two of the connection weights, and these two signals are added together as part of the unit's net input to the pattern. The bias of the unit (i.e., the μ value of the unit's Gaussian activation function) is then added to the two signals to complete the net input calculation. If net input is near zero, the unit will convert this into a high level of internal activity. If net input is too far above or too far below zero, then the unit will convert this into a low level of internal activity.

Because the weights for a hidden unit represent the heading of an individual city, when net input is computed in this way it must represent a combination of two headings. That is, the hidden unit is not simply measuring the angular displacement between two cities. Instead, it measures the bearing of one city, and the bearing of the other city, where both bearings are taken with respect to the hidden unit's heading. It then combines the two bearings together, either by adding them or by subtracting them.

To explore how the hidden units were combining bearings, for each hidden unit we created a 13×13 net input matrix. Each row and column in this matrix corresponded to a city, and each entry in the matrix was the net input for the unit for a particular pair of cities. We also created a 13×13 bearing matrix for each hidden unit. Every entry in this matrix was the cosine of the bearing of the row city (taken with respect to the unit's heading) combined with the cosine of the bearing of the column city. For some matrices, we combined the bearings by adding them together. For other matrices, we combined the bearings by taking their difference. The particular method of combination that we selected was the one that led to the highest correlation between the two matrices.

Table 4 presents the correlations between the net input matrix and the bearing matrix for each hidden unit. Once again the correlations between a property derived from the map of Alberta (i.e., combined bearings) and a property of the network are strong for each hidden unit. Indeed, the correlations in this table are very similar to those reported Table 3, and F-tests again show that all are statistically significant.

Table 4
Correlation between net input matrices and bearing matrices for each hidden unit

Hidden Unit	Correlation	Operation
H1	0.61	Bearings are added
H2	-0.73	Bearings are added
H3	0.54	Bearings are subtracted
H4	0.87	Bearings are added
H5	0.59	Bearings are subtracted
H6	0.72	Bearings are added
H7	0.75	Bearings are subtracted

On the one hand, the fact that the two sets of correlations are similar should not be too surprising. The first set of correlations involves weight spaces, which are computed from the connection weights. Connection weights are also a key component of the calculation of the net inputs that are involved in the second set of correlations. Nevertheless, the fact that the correlations in Table 4 are also high provides important converging evidence to support the claim that the hidden units are sensitive to direction. This is because weight space, used to calculate the Table 3 correlations, is a description of hidden unit structure that is an abstraction over the actual connection weights. Net input is much more concretely related to hidden unit operations, and the fact that correlations involving it remain high is encouraging.

Heading and Hidden Unit Activation

When a stimulus pattern is presented to the network, it causes a signal to be sent to the network's hidden units. The hidden units then use this signal to compute their net input, as discussed in the previous section. The net inputs are then converted into a number between 0 and 1 that represents each hidden unit's internal level of activity to the incoming signal. This is accomplished by applying the Gaussian equation to the net input.

The use of the Gaussian activation function in value units means that each hidden unit is tuned to respond to a relatively narrow range of net inputs (e.g., Dawson, 2004, chap. 10). In the previous section, it was shown that net inputs represented a combination of two city bearings relative to a hidden unit's heading. Therefore the hidden value units in this network are tuned to respond to a relatively narrow set of city bearing combinations.

To illustrate this point, let us consider Hidden Unit 6. As reported in Table 4, this unit operates by adding two city bearings (i.e., the cosines of the bearings as encoded in connection weights) to a bias value of 0.02 to compute net input. When net input is equal to 0, the unit will generate a maximum activation of 1. What combinations of city bearings will lead to such high activation for this unit? What combinations of city bearings will result in the unit generating little or no activation?

One can use a unit's weight space to make predictions about unit activation. Figure 4 is a graph of the weight space for Hidden Unit 6. The vector in Figure 4 represents the heading of the hidden unit, and the origin of this vector is also the origin of the weight space. The line is the weight space, and the location of each city on this line is taken from Hidden Unit 6's column in Table 2. As noted in the earlier discussion of Table 3, the location of each city on this line also serves as a representation of the city's bearing relative to the unit's heading. Because of the way in which connection weights are projected onto the weight space, two properties must be true in order for the net input for this unit to be near zero. First, the two cities must be located on the same side of the heading vector (that is, both must have positive coordinates, or both must have negative coordinates). Second, the two cities must be located roughly the same distance from the origin

of the weight space. When these two conditions are true, the signals being sent through the connection weights from the two cities will cancel each other out, resulting in near zero net input, and near maximum activation.

Keeping these two properties in mind, we would predict that Hidden Unit 6 should generate high activity when Lloydminster and Fort McMurray are presented to the network, as well as when Camrose and Edmonton are presented. This is because the two required properties are true for both of these city pairs. In fact, the unit generates an activation of 1.00 to both of these pairs. We predict that this unit will generate intermediate activity when Banff and Medicine Hat are presented to the network. This is because while the two cities are on the same side of the origin, they are at different distances from it. Hidden Unit 6 has an activation of 0.46 to this pair. We also predict that when Edmonton and Fort McMurray are presented to the network, Hidden Unit 6 will generate even weaker activity. This is because even though both cities are roughly the same distance from the origin, they are on opposite sides of it. Hidden Unit 6 produces an activation of 0.11 to this city pair.

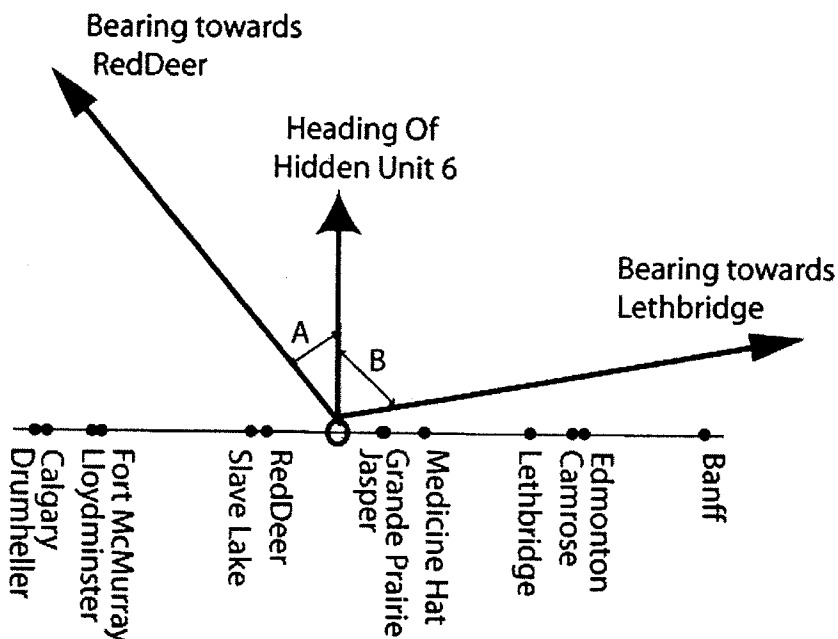


Figure 4. A graphical interpretation of the operation of hidden units in the network, using Hidden Unit 6 as an example. The horizontal line represents the positions of the thirteen cities on the unit's weight space. The circle represents the origin of this space. The unit's heading is given from this origin. Angle A is the bearing towards RedDeer relative to this heading, and Angle B is the bearing towards Lethbridge. The position of a city on the weight space is highly correlated with the cosine of the city's bearing.

There is another graphical approach for rapidly evaluating each hidden unit's sensitivity to combinations of city bearings. This approach is only possible because the hidden units have been shown to convert a categorical input variable (i.e., city name) into a value on a continuous scale (i.e., city bearing). As a result, we can illustrate the receptive field of each hidden unit by using bubble plots to graph hidden unit activity in a two-dimensional space, where one dimension is the bearing of the *from city* and the other dimension is the bearing of the *to city*. Figure 5 presents the receptive fields of each of the hidden units.

From Figure 5, it is clear that all of the hidden units are tuned to respond to combinations of city bearings that fall into a particular range. However, some of the hidden units are more finely tuned than others. For example, Hidden Unit 2 only generates responses when both the *from bearing* and the *to bearing* are in the range from 0° to 62° . Furthermore, this unit typically generates a maximum response when the two bearings are equal. Similarly, Hidden Unit 4 only responds when the *from bearing* and the *to bearing* are both in the range from 120° to 176° . Again, for stimuli that fall into this narrow range, it tends to generate higher responses when both bearings have similar values, and to generate lower responses when the two bearings have dissimilar values. This kind of tuning is to be expected because the hidden units use a Gaussian activation function, which of course is tuned to respond to a narrow range of net inputs. The fact that all of the receptive fields fall into a range that never goes beyond 180° is also due to the Gaussian activation function. The symmetry of this function means that negative net inputs and positive net inputs are treated identically.

While all of the hidden units exhibit directional tuning, some appear to be more broadly tuned than others. This is possible because while some of the tuning of the reception field depends on the Gaussian activation function, some of the tuning also depends on the particular viewpoint that a hidden unit has on the cities (i.e., a hidden unit's location and heading, as reported in Table 3). Hidden Unit 3 provides one example of a more broadly tuned unit, because it generates responses when both bearings are in the range from 84° to 168° , but also generates a few responses when the *from bearing* is at 0° . Similarly, Hidden Unit 1 responds to stimuli for which both bearings are in the range from 24° to 177° . As well, Hidden Unit 6 responds when both bearings are in the range from 9° to 176° .

One notable feature of the tuning ranges that are reported above, and which can be observed in Figure 5 (i.e., by comparing the arc swept by the data as it rotates in the direction of the *from heading axis* to the width occupied by the data in the direction of the *to heading axis*), is that the *from bearings* that a unit is sensitive to fall into the same range as the *to bearings* that a unit is sensitive to. One reason for this would be if the receptive fields are symmetric, which would mean that a hidden unit would generate the same response to a pair of cities regardless of which was the *from city* and which was the *to city*.

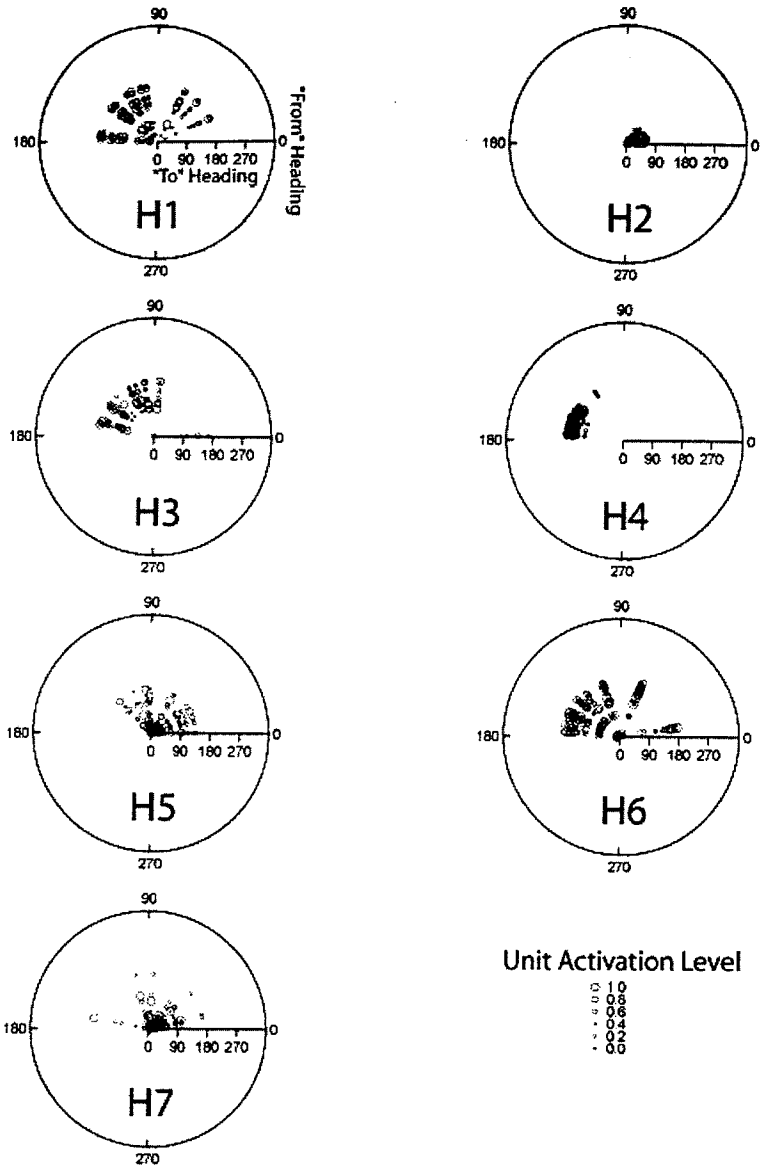


Figure 5. Bubble plots of the activity of each hidden unit as a function of the bearings of two cities (a *from* city and a *to* city) for each pair of cities in the training set. The size of the bubble reflects the degree of hidden unit activity. The coordinates of the bubble indicate the bearings of the two cities. The circular axis provides the bearing of the *from* city, and the horizontal axis provides the bearing of the *to* city.

To determine the degree of symmetry in these responses, we created an activation matrix for each hidden unit. This was done by taking each unit's net input matrix (described earlier), and by using the Gaussian activation function to convert each net input into an activation value. We then computed the symmetric portion of each matrix by averaging it across the diagonal; that is, by replacing the value at row x and column y , and replacing the value at row y and column x , with the average of the two values (e.g., Dawson & Harshman, 1986). To determine the proportion of variance in the activation matrix that was symmetric, for each hidden unit we took the correlation between the full activation matrix and its symmetric component, and then squared the correlation. The squared correlations ranged from a low of 0.974 (for Hidden Unit 1) to a high of 1.00 (for Hidden Units 1, 2, 6, and 7). In other words, every hidden unit had a nearly perfectly symmetric receptive field. On the one hand, this property is not completely surprising, because the responses that the network was trained to make in its output units were symmetric. On the other hand, there is no a priori restriction that such symmetry must also be true of the units that make up the network's internal structure, and so discovering it is interesting.

Combining Hidden Unit Responses to Generate Output Ratings

The results that have been described to this point have shown that each hidden unit responds to a combination of a pair of city bearings, where these bearings are dependent upon each hidden unit's position and heading on the map of Alberta. After a pair of city names has been presented to the network, the output units take hidden unit responses and combine them to make a distance rating response.

A general idea of how hidden unit responses are being used to make distance judgments can be obtained by eliminating the output layer of the network altogether, and by using hidden unit activities to each pattern to predict the distance rating. This approach is less powerful than the actual network, because it replaces the nonlinear activity of the output layer with a linear statistical method that can at best approximate network behavior. However, the use of traditional statistics to predict distance ratings from hidden unit activities can provide some insight into the process of converting directions into distances.

Table 5 presents the regression coefficients (and the associated statistical tests) when the activities of the 7 hidden units are used as predictors of distance ratings in a multiple regression equation. The R^2 for this regression equation was 0.72, indicating that a linear combination of hidden unit activities provides an excellent predictor of distance ratings, even when these activities are not subject to further nonlinear processing, as is actually the case in the trained network.

An examination of Table 5 indicates that as far as the regression equation is concerned, four of the hidden units (Hidden Unit 1, 2, 5, and 6) are all roughly equal contributors to the prediction of the distance rating, are all given fairly large regression coefficients, and are all statistically significant predictors. The other three hidden units are not statistically significant predictors, and are given much smaller regression coefficients. It is likely that these hidden units are far

Table 5

Regression coefficients, t statistics, and probability levels for each hidden unit when hidden unit activity is used in a regression equation to predict distance rating. The correlation in the final column is that between a hidden unit's activation and the distance rating

Hidden Unit	Coefficient	t	p(2-tail)	Correlation With Distance Rating
H1	2.572	4.417	0.000	0.007
H2	1.947	2.687	0.008	0.217
H3	-0.388	-0.581	0.562	-0.147
H4	-0.395	0.784	0.434	-0.541
H5	2.514	3.559	0.000	0.271
H6	2.722	4.437	0.000	-0.088
H7	0.696	1.195	0.234	-0.212

more important when involved in the nonlinear transformations performed by the network, which can account for approximately 28% more of the variance in distance ratings than can this linear regression equation.

Table 5 also provides the simple correlation between a hidden unit's activations and the distance ratings. These correlations are fairly small, indicating that individual hidden units are by themselves not very accurate predictors of distance ratings. More accurate prediction requires using a combination of hidden unit responses. This is consistent with the notion that the internal structure of this particular network uses coarse coding, as well as the argument made earlier that estimates of distance based on a single set of angular measurements will be intrinsically inaccurate.

Discussion

When artificial neural networks are incorporated into synthetic psychology, three general research steps are required (Dawson, 2004). First, a network is synthesized, usually by training it to perform a desired mapping between input stimuli and output responses. Second, the internal structure of the trained network is analyzed, in an attempt to discover what regularities are being detecting to perform this mapping, and how the network represents these regularities. Third, emergent properties are sought in the results of this analysis. In particular, the question is whether the network has discovered new, surprising, or interesting ways of solving the problem.

The network that has been described in the preceding sections of this paper serves as an example of a case study in synthetic psychology. With respect to the first step, synthesis, this network was trained to make ratings of the crow's flight distance between all possible pairs of thirteen Albertan cities. The bulk of the results section was devoted to the second step, the analysis of this network. To our surprise, there was no evidence that the hidden units of this network

represented distances at all, in spite of the fact that the network was being trained to make distance judgments, and in spite of the fact that a previous network trained on this problem was sensitive to distance (Dawson et al., 2000). Instead, the evidence indicated that the hidden units represented directions, and used measurements of city bearings to derive distance ratings.

In particular, it was found that each hidden unit could be located at a particular latitude and longitude on the map of Alberta. Furthermore, at this location, each hidden unit had associated with it a specific heading. Each hidden unit had a unique position and heading, providing seven different perspectives from which the thirteen cities could be viewed. A hidden unit could be interpreted as measuring the bearing of each city relative to the hidden unit's heading. In essence, the connection weight from a city to the hidden unit encoded this bearing; connection weights were strongly correlated with the cosines of bearings for all of the hidden units.

Bearings were used to compute distance ratings as follows: When a stimulus was presented to the network, two city input units were activated, sending a pair of signals to each hidden unit. Each pair of signals represented the bearing of the two cities relative to a hidden unit heading. Each hidden unit combined the two bearings (some by adding them, some by subtracting them) and then used the Gaussian activation function to convert the combined headings into an activity level in the range from 0 to 1. The network could generate an accurate distance rating from a weighted combination of the hidden unit activities, after performing a second Gaussian transformation.

The third step in using artificial neural networks in synthetic psychology is to explore the emergent properties of the analyzed network. This is carried out in the remaining sections of this paper, which show that this directional representational scheme has several interesting properties that may be relevant to researchers interested in the neural encoding of cognitive maps.

General Characteristics of the Network's Representation

One of the general issues that motivated the current research was the potential for difference between the properties of a representation and the properties of the domain that it represents. We argued earlier that the assumption that a representation must have map-like properties to serve as a cognitive map was an example of committing the scope slip. The current network could be viewed as internalizing the map of Alberta in the sense that it could make accurate distance judgments between any pair of cities that it was trained on. However, a brief consideration of its general characteristics suggests that the network is not itself particularly map-like in nature.

First, the network uses coarse coding. One characteristic of coarse coding is that the local measurements that it uses are not particularly accurate, and this is certainly true of our network. While the correlations between map properties (i.e., cosines of city bearings) and connection weights are very strong, they are certainly not perfect, indicating a degree of the inaccuracy. This is an example of a lack of isomorphism between the network and the map that it represents.

This kind of lack of isomorphism has been used by previous researchers to criticize the view that place cells in the hippocampus mediate a cognitive map because they are at best locally metric (i.e., they are relatively inaccurate).

Second, in coarse coding, inaccurate local measurements are compensated by combining several of them together. The logic of this approach is that if each local measurement has a different source of random error, then this error will tend to cancel its self out when the measurements are combined, resulting in a stronger signal to noise ratio. The network's representation exhibits this property. First, the output units each compute a weighted combination of the signals coming for all seven hidden units. Second, each hidden unit has a different perspective (i.e., a different location and a different bearing) on the map, resulting in a different source of error.

In neuroscience, a coarse coding scheme is sometimes called an ensemble code. Such codes have been proposed in accounts of animal navigation. Kamil and Cheng (2001) have proposed a multiple bearings theory in which birds localize food sites by representing the bearings of the site to multiple landmarks. The use of multiple landmarks increases accuracy in their theory. Similarly, Jensen and Lisman (2000) have shown that ensemble coding of hippocampal place cells can provide a more accurate measurement of an animal's position in space than can the output from a single place cell. However, the representation of space with an ensemble code is far from map-like, because there is no obvious relationship between the properties of this type of code and those of a traditional map.

Third, it is usually assumed that if a brain-like system is serving as a map, then its components should be topographically organized. The lack of topographical organization of place cells has been used to criticize the claim that the hippocampus serves as a cognitive map. It is less clear in the current network, whose components are sensitive to direction instead of location, what a topographic organization would entail. However, one might expect that neighboring hidden units would have similar bearings in a topographic organization. This is not the case for the hidden units of the network, indicating again that its general characteristics are not map-like.

Previous Research on Sensitivity to Direction

To many, it may not be surprising that the hidden units of the artificial neural network were sensitive to direction. First, for any map to be effective it must encode direction as well as distance. Second, there is a wealth of behavioral and physiological evidence that many animals are sensitive to direction. This includes results that show some animals are capable of measuring the angle subtended between two landmarks. Some of this evidence is briefly reviewed below.

One source of behavioral evidence that shows that navigating animals are sensitive to angles involving landmarks comes from the study of bees. Research on bee navigation is controversial in the sense that there is no universal agreement as to whether bees use cognitive maps or not. Early results seemed to

indicate that displaced bees were capable of determining a novel shortcut to a food source (Gould, 1986). However, this result has been difficult to replicate (e.g., Dyer, 1991; Wehner, Bleuler, Nievergelt, & Shah, 1990; Wehner & Menzel, 1990). Because of this, some researchers are of the opinion that bees do not have cognitive maps but are instead route following agents (but for an alternate position, see Gould, 2002). For example, Cheng's (2000) theory of bee navigation involves four different servomechanisms that are activated in sequence. First, bees follow a route to a general food region. Second, bees get near a food source in this region by using landmarks. Third, a sensorimotor vector to target can be initiated (optionally) on the basis of visual information to take the bee directly to a food source. Fourth, if the bee cannot vector to target, then it moves about in a search pattern in order to match landmarks to positions on its retinae to determine the precise location of a remembered food source.

While there is considerable debate about whether bees use cognitive maps, there is much more agreement about the kind of variables that bees used to navigate. Of interest in the current manuscript is the fact that bees appear to be sensitive to both allocentric and egocentric angles involving landmarks.

For example, consider Cheng's (2000) first servomechanism, following a route. The famous tail-wagging dance is used by one bee to communicate the direction of the route to other bees in the hive (e.g., von Frisch, 1966, 1967). The tail-wagging dance is approximately in the shape of a figure eight, and the waggle component is the middle of this figure. When the dance is performed inside the beehive, the angle of the waggle component relative to the gravitational axis of the hive specifies the direction of a food source. In particular, the direction that is specified is the angle between the food source and the azimuth of the sun, with the beehive at the origin of the angle. Thus, bees are capable of measuring this angle, communicating this angle to other bees, and using this angle to determine a route direction that can be followed using other mechanisms, such as the sun compass.

The tail-wagging dance communicates an allocentric angle that is based upon three different locations. It has also shown that bees are sensitive to egocentric angles, which is probably important to the fourth servomechanism in Cheng's (2000) theory. Consider a study conducted by Cartwright and Collett (1983). In one experiment, a bee was trained to feed at a food source located at a particular position relative to a single landmark, which was a black cylinder. In some test conditions, the size of the cylinder was changed when the test bee was absent from the testing room. The results suggested that the bee was using the apparent size of the landmark as a distance cue. This is because when the cylinder was enlarged the bee searched for food at a location that was further away from the landmark, and when the cylinder was made smaller the bee search for food at a location that was closer to the landmark.

However, when an array of three cylindrical landmarks was used as a location cue, the bee's behavior was not affected by manipulating landmark size. Instead, Cartwright and Collett (1983) found that search behavior could be explained using a simple directional rule. When the bee feeds at the training site, each of

the three landmarks have a particular compass bearing on its retina relative to the food location. "During tests the bees would search where the compass bearings of the landmarks on its retina match those experienced at the food source during training" (p. 529). Interestingly, a computer simulation that provided a good fit to bee behavior represented the retinal snapshot of the landmarks using Earth-bound coordinates instead of retinal coordinates.

The results obtained by Cartwright and Collett (1983) clearly indicate that bees are sensitive to the bearings of an array of landmarks in locating a food source. Other experimental results have demonstrated a similar sensitivity in vertebrates.

For example, several studies have shown that birds can learn a variety of geometric rules involving the relationship of landmarks, including rules involving directions or angles. For example, Clark's nutcrackers can learn to find a location that lies at the apex of a triangle when the position of the apex is specified solely by the angle subtended by two landmarks (Kamil & Jones, 2000). These birds can also learn the position of location that is defined by a constant angle with respect to two landmarks, whether this angle is absolute (i.e., the landmark configuration has a constant orientation relative to the walls of a testing room) or relative (i.e., the landmark configuration rotates in the room from trial to trial), though use of absolute bearings appears to be easier than the use of relative bearings (Jones & Kamil, 2001). Kamil and Jones (1997) also found that these birds could be trained to find a location defined as the midway point between two landmarks, and that part of this ability required the detection of directional bearings. Importantly, the use of directional information was more accurate than the use of distance information in performing this task. Kamil and Cheng (2001) developed a model of the nutcracker's abilities to learn such rules that was based on the fundamental assumption that these birds represent directional bearings to multiple landmarks. Spetch, Rust, Kamil, and Jones (2003) demonstrated that pigeons were also capable of learning geometric rules that depended upon a sensitivity to the angle between two landmarks, but that this ability was more modest than that demonstrated by Clark's nutcrackers.

There is also evidence that directional information is encoded in the rat hippocampus, as well as other regions in the rat's brain. In addition to the place cells that were described in the introduction, researchers have also found neurons that they have called head direction cells (for overviews, see Redish, 1999; Sharp, Blair, & Cho, 2001; Taube, 1998). In general, head direction cells have a preferred direction of firing, and generate strong signals when a rat's head is pointing in this preferred direction. "In this manner, cell activity can be viewed as analogous to the response of a compass needle in a local environment, which always points north no matter where one is located" (Taube, 1998, p. 226). Usually, head direction neurons for a variety of different directions are available, and cooperative and competitive interactions between these different cells result in one direction being selected (and represented with neural activity in the appropriate cells). Most models of this type of processing use an attractor network (a self-organizing artificial neural network) with different head direction

cells linked with excitatory and inhibitory connections to create a directional cell assembly (e.g., Redish, 1999).

The source of this directional sensitivity is complex. While the responses of head direction cells can come under the control of visual landmarks, these neurons are also able to maintain their directional heading when such landmarks are absent, most likely under the control of self-motion cues such as vestibular input (e.g., Knierim, Kudrimoti, & McNaughton, 1998). Such idiothetic cues as this latter type are an important source of input to head direction cells, and as a result these cells are well suited for performing navigation via path integration (e.g., Gallistel, 1990). There are many unanswered questions about head direction cells, including how their signals are used to guide an animal's behavior, and how self-motion and landmark cues are integrated to control their response (Taube, 1998).

In many respects, the kind of information that head direction cells appear to encode differs from the kind of information to which the hidden units in the artificial neural network are sensitive. For instance, head direction cells seem to capture a unary feature (e.g., the compass direction of the animal's head) while the hidden units capture a binary feature (e.g., the angle between a unit's bearing and a city on the map). Nevertheless, head direction cells do show that the neural encoding of some directional information is both possible and important.

A Novel Contribution: Using Direction to Compute Distance

From the literature reviewed in the previous section, it is clear that the kind of directional sensitivity found in the artificial neural network is not unique. Many different animals exhibit this sort of sensitivity, although the relationship between the neural encoding of direction and the kind of encoding discovered in the network is an open question. However, there is one aspect of the artificial neural network that is very surprising, and which could serve as a springboard for a different kind of research on animal navigation. This surprising characteristic is the fact that the artificial neural network is using directional measures to mediate judgments about distance.

In cognitive science, researchers often adopt what is known as the modularity hypothesis (Fodor, 1983). According to this hypothesis, there is a variety of extremely different information processing problems that must be solved by cognition. A specific algorithm is required to solve a particular problem, and in many cases the algorithm for solving one problem will not be suited to solve another. Because of this, it is extremely unlikely that a single, large, general-purpose method is used to solve all these problems. Instead, it is assumed that each problem is solved by a small, specialized procedure, and each procedure is devoted to a particular type of computation. It is often assumed as well that each procedure is associated with specific neural circuitry.

The modularity hypothesis is also prevalent in the study of animal navigation (e.g., Shettleworth, 2002). The reason for this is that different components of navigation present different computational problems that require specialized solutions. For example, we have previously mentioned that Redish (1999) has

proposed a navigational theory that includes five separate modules, each of which is associated with a particular navigational task and with characteristic neural circuitry. Importantly, problems involving distance and problems involving direction often appear to be associated with different modules in theories of animal navigation. In most theories, the mechanisms that compute location, position or distance are usually proposed as being quite distinct from those that compute direction or angle.

This distinction is evident in observations about the functional role of different neurons in the hippocampus. Place cells are usually viewed as representing an animal's location and are involved in representing the distance from one position to another (e.g., O'Keefe & Nadel, 1978). They are usually not described as being directionally sensitive in theories of the cognitive map, in spite of the fact that some studies have observed that place cells may in some instances be sensitive to the direction in which an animal is moving (e.g., McNaughton, Barnes, & O'Keefe, 1983; Muller, Bostock, Taube, & Kubie, 1994). Instead, directional sensitivity is ascribed in hippocampal theories to a separate locus, the head direction cells (e.g., Sharp et al., 2001; Taube, 1998). Furthermore, head direction cells are frequently proposed to be involved in dead reckoning computations, which can be thought of as being a different kind of navigation than the sort that is mediated by place cells (e.g., Redish, 1999).

Recent theories of bird navigation also distinguish between modules concerned with direction and modules concerned with distance or position. For example, Jacobs (2003) and Jacobs and Schenk (2003) have recently proposed a theory of navigation, the parallel map theory, which attempts to build relationships with theories derived from the study of the rat hippocampus. In the parallel map theory, there are two major subsystems. The first of these is primarily directional, and is called the *bearing map*. The bearing map is a two-dimensional coordinate system constructed from the integration of self-motion and directional cues, such as sensory gradients. It can be used to navigate over very long distances, and permits an animal to maintain an accurate representation of position even in unknown territory. The second component of the parallel map theory is the *sketch map*. It is a representation of specific landmarks that can be described as a fine-grained topographic map of a small, local area. The sketch map represents positional cues, and is therefore similar in nature to the cognitive map described by O'Keefe and Nadel (1978). A final component of the parallel map theory is that at a higher level the bearing map and the sketch map can be combined into a single representation, the *integrated map*, which can be used to compute novel routes among set of landmarks.

The parallel map theory is an important example of the modularization of direction and distance. The bearing map represents direction, while the sketch map represents position. These modular distinctions in the parallel map theory are also evident in other theories of the navigation. For example, each of the servomechanisms in Cheng's (2000) theory can be thought of as being modules. The first servomechanism in Cheng's theory, route following, is functionally

similar to Jacob and Schenk's (2003) bearing map. Later servomechanisms, in particular image matching, are more related to the sketch map.

Why is the modular distinction between direction and distance so prevalent in theories of animal navigation? One reason is that, at first glance, it seems very difficult to use direction information to compute distance, or vice versa. In other words, it seems that direction and distance should be associated with different modules because they appear to be computationally distinct. This can be demonstrated by determining the extent to which a rich set of directional measurements can be used to compute the distances between locations.

We created a table of pure directional measures relating the thirteen cities on the map of Alberta. This was accomplished by taking each city in turn as occupying the origin of the map, and assigning to this origin a unit vector pointing north. We then computed the cosine between this vector and the vectors from the origin to all of the cities on the map. This resulted in a thirteen by thirteen matrix of cosines. Multidimensional scaling (e.g., Kruskal & Wish, 1978) was then conducted on this matrix to convert the cosines into a map-like configuration. It was found that a two dimensional map, derived from the cosines, accounted for 99.9% of the variance in the data.

The question of interest was the relation between the distances between cities in the multidimensional scaling solution—which were based purely on direction—to the actual distances between cities on the map of Alberta. We correlated the two different thirteen by thirteen distance matrices with each other, and found that the correlation was only 0.69. In other words, a rich set of pure directional measurements could only be used to account for 47.61% of the variance in the actual distance data. With a result like this, it is not surprising that directional and positional processes are assigned to different modules in many theories of animal navigation.

Such a result makes the general nature of the artificial neural network truly remarkable. All of the evidence found in the analysis of that network indicated that the hidden units were making very specific types of directional measurements. In turn, the network used these directional measurements to make accurate distance judgments. The network's key contribution to the theory of cognitive maps is in providing an example of the use of direction to compute distance, a contribution that is extremely novel.

One of the main implications of this contribution is that the network provides a plausible reason for reevaluating the traditional modularization of direction and distance in theories of animal navigation. This is important because there are many unanswered questions about the cognitive map, and it is useful to not prematurely rule out possible avenues for exploring their answers.

For instance, many researchers are comfortable with the proposal that place cells represent locations in the environment. Nevertheless, the information that determines place cell firing it is far from clear. Reliable place fields can be established in the absence of visual information, and place cell activity can also be affected by intrahead cues, such as idiothetic information, memory, and motivational state (Best et al., 2001). This broad range of sensitivities has lent

credence to claims that place cells are not the primitives of a cognitive map, but have more general roles (e.g., Eichenbaum, 2002; Eichenbaum et al., 1999). In spite of such claims, there is also evidence that place cells are sensitive to information that is pertinent to a spatial role, and this evidence is relevant to the merging of directional and distance information that is suggested by our network interpretation.

Some evidence suggests that place cells are sensitive to the distances to two or more landmarks (O'Keefe & Burgess, 1996). O'Keefe and Burgess measured place fields in four different walled chambers: a small square, a large square, a vertical rectangle, and a horizontal rectangle. Place fields changed from chamber to chamber, but in a systematic way that could be related to the distance from the animal to each of the four walls in a chamber. O'Keefe and Burgess were able to model the place fields as the thresholded sum of the distance to each wall, after the distance had been transformed by a Gaussian function. This model was able to explain changes in the shapes of place fields when chamber walls were moved to different positions (e.g., changes from the small square chamber to the large square chamber).

In the O'Keefe and Burgess (1996) study, the chamber walls served as landmarks for the place cells, and were in a fixed orientation. They then manipulated the relative distances between the landmarks. Fenton, Csizmadia, and Muller (2000a) adopted a different experimental approach, in which the distances between landmarks and the center of the chamber were held constant. In their study, place fields were measured when animals were in a cylindrical chamber. A black and a white cue card were placed on the wall of this chamber to serve as prominent landmarks. When both cards were rotated to different positions on the chamber, maintaining the relative positions of the cards, the place fields rotated accordingly. Removal of one card, and the rotation of the other, again led to uniform rotations of the place fields. However, when the angular distance between the two cards was changed, by either moving the cards closer together on the chamber wall, or by moving them apart, the shape of the place fields became distorted, the local smoothness of the place fields decreased, and the firing rates of the place cells decreased.

Fenton, Csizmadia, and Muller (2000b) used a vector field equation to model the changes of the place fields under all of their experimental manipulations. They identified two separate transformations of each component of a place field, a rotation and a translation. Both transformations were based upon a small set of measurements: vectors representing the change in position of each cue card from one condition to the next, the distance of the field center from each cue card, the radius of the cylinder, and the distance of the field from the center of the cylinder. The vector field equation that used these measurements was able to provide very good fits to most of the experimental results, in some cases accounting for as much as 88% of the variance in the data.

Fenton et al. (2000b) point out several differences between their model and the one proposed by O'Keefe and Burgess (1996). However, both models are similar in that they are both primarily distance based: most of the variables in

both models involve distance measurements between place cells and landmarks, or between place cells and chamber positions (e.g., the center of the cylindrical chamber). One exception to this are the two vectors that Fenton, Csizmadia and Muller used to represent the change in position of the cue cards, which include information about distance as well as direction. For instance, these vectors are highly correlated with the angle of rotation of the cards with respect to the center of the cylinder.

The network described in the current paper suggests one plausible line of inquiry concerning alternative vector field models of place cell activity. Rather than using primarily distance-based measures, it is possible that vector fields might be defined using equations that take as input vectors that relate the position of place cells to environmental markers. These vectors would change length under manipulations like those used by O'Keefe and Burgess (1996), and would change both orientation and length under manipulations like those used by Fenton et al. (2000a). In terms of the modularity issue raised in this section, such vectors would naturally combine distance and direction information.

The network that we examined in this paper raises this possibility because it has demonstrated that directional information can be converted into positional information (e.g., distance between landmarks), and therefore the two types of information need not be modularized. Furthermore, robotics researchers are aware that directional measurements are far easier to make than are distance measurements, and have shown that angle-based representational schemes can be used for robot navigation (e.g., Bekris, Argyros, & Kavraki, 2004). In other words, one general issue that is raised by the results that we have described above is the extent to which place cells are sensitive to directional or angular information, in addition to their sensitivity to distance.

Acknowledgments

This research was supported by NSERC and SSHRC grants awarded to MRWD. Many thanks to Dr. Clayton Dickson and Dr. Marcia Spetch of the psychology department at the University of Alberta for their helpful suggestions.

References

- Ballard, D. (1986). Cortical structures and parallel processing: Structure and function. *Behavioral and Brain Sciences*, 9, 67–120.
- Bekris, K. E., Argyros, A. A., & Kavraki, L. E. (2004). Angle-based methods for mobile robot navigation: Reaching the entire plane. In *Proceedings of the 2004 IEEE International Conference on Robotics & Automation* (pp. 2373–2378). New Orleans, LA.
- Bennett, A. T. D. (1996). Do animals have cognitive maps? *Journal of Experimental Biology*, 199, 219–224.

- Berkeley, I. S. N., Dawson, M. R. W., Medler, D. A., Schopflocher, D. P., & Hornsby, L. (1995). Density plots of hidden value unit activations reveal interpretable bands. *Connection Science*, 7, 167–186.
- Best, P. J., White, A. M., & Minai, A. (2001). Spatial processing in the brain: The activity of hippocampal place cells. *Annual Review of Neuroscience*, 24, 459–486.
- Braitenberg, V. (1984). *Vehicles: Explorations in synthetic psychology*. Cambridge, MA: MIT Press.
- Burgess, N., Recce, M., & O'Keefe, J. (1995). Spatial models of the hippocampus. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks* (pp. 468–472). Cambridge, MA: MIT Press.
- Cartwright, B. A., & Collett, T. S. (1983). Landmark learning in bees: Experiments and models. *Journal of Comparative Physiology*, 151, 521–543.
- Cheng, K. (2000). How honeybees find a place: Lessons from a simple mind. *Animal Learning & Behavior*, 28, 1–15.
- Cheng, K., & Spetch, M. L. (1998). Mechanisms of landmark use in mammals and birds. In S. Healy (Ed.), *Spatial representation in animals* (pp. 1–17). Oxford: Oxford University Press.
- Dawson, M. R. W. (1998). *Understanding cognitive science*. Oxford, UK: Blackwell.
- Dawson, M. R. W. (2004). *Minds and machines: Connectionism and psychological modeling*. Malden, MA: Blackwell.
- Dawson, M. R. W., Boechler, P. M., & Valsangkar-Smyth, M. (2000). Representing space in a PDP network: Coarse allocentric coding can mediate metric and nonmetric spatial judgements. *Spatial Cognition and Computation*, 2, 181–218.
- Dawson, M. R. W., & Harshman, R. A. (1986). The multidimensional analysis of asymmetries in alphabetic confusion matrices: Evidence for global-to-local and local-to-global processing. *Perception & Psychophysics*, 40, 370–383.
- Dawson, M. R. W., Medler, D. A., & Berkeley, I. S. N. (1997). PDP networks can provide models that are not mere implementations of classical theories. *Philosophical Psychology*, 10, 25–40.
- Dawson, M. R. W., Medler, D. A., McCaughan, D. B., Willson, L., & Carbonaro, M. (2000). Using extra output learning to insert a symbolic theory into a connectionist network. *Minds and Machines*, 10, 171–201.
- Dawson, M. R. W., & Piercey, C. D. (2001). On the subsymbolic nature of a PDP architecture that uses a nonmonotonic activation function. *Minds and Machines*, 11, 197–218.
- Dawson, M. R. W., & Schopflocher, D. P. (1992). Modifying the generalized delta rule to train networks of nonmonotonic processors for pattern classification. *Connection Science*, 4, 19–31.
- Dawson, M. R. W., & Zimmerman, C. (2003). Interpreting the internal structure of a connectionist model of the balance scale task. *Brain & Mind*, 4, 129–149.
- Duch, W., & Jankowski, N. (1999). Survey of neural transfer functions. *Neural Computing Surveys*, 2, 163–212.

- Dyer, F. C. (1991). Bees acquire route-based memories but not cognitive maps in a familiar landscape. *Animal Behaviour*, *41*, 239–246.
- Eichenbaum, H. (2002). *The cognitive neuroscience of memory: An introduction*. Oxford ; New York: Oxford University Press.
- Eichenbaum, H., Dudchenko, P., Wood, E., Shapiro, M., & Tanila, H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, *23*, 209–226.
- Fenton, A. A., Csizmadia, G., & Muller, R. U. (2000a). Conjoint control of hippocampal place cell firing by two visual stimuli I: The effects of moving the stimuli on firing-field positions. *Journal of General Physiology*, *116*, 191–209.
- Fenton, A. A., Csizmadia, G., & Muller, R. U. (2000b). Conjoint control of hippocampal place cell firing by two visual stimuli II: A vector-field theory that predicts modifications of the representation of the environment. *Journal of General Physiology*, *116*, 211–221.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Gallistel, C.R. (1998). Insect navigation: Brains as symbol-processing organs. In D. Scarborough & S. Sternberg (Eds), *An invitation to Cognitive Science: Vol. 4. Methods, models and conceptual issues* (2nd ed., pp. 1–52). Cambridge, MA: MIT Press.
- Gould, J. L. (1986). The locale map of honey-bees: Do insects have cognitive maps? *Science*, *232*, 861–863.
- Gould, J. L. (2002). Can honey bees create cognitive maps? In M. Bekoff, C. Allen & G. M. Burghardt (Eds.), *The cognitive animal* (pp. 41–45). Cambridge, MA: MIT Press.
- Jacobs, L. F. (2003). The evolution of the cognitive map. *Brain, Behavior and Evolution*, *62*, 128–139.
- Jacobs, L. F., & Schenk, F. (2003). Unpacking the cognitive map: The parallel map theory of hippocampal function. *Psychological Review*, *110*, 285–315.
- Jensen, O., & Lisman, J. E. (2000). Position reconstruction from an ensemble of hippocampal place cells: Contribution of theta phase coding. *Journal of Neurophysiology*, *83*, 2602–2609.
- Jones, J. E., & Kamil, A. C. (2001). The use of relative and absolute bearings by Clark's nutcrackers, *Nucifraga columbiana*. *Animal Learning & Behavior*, *29*, 120–132.
- Kamil, A. C., & Cheng, K. (2001). Way-finding and landmarks: The multiple-bearings hypothesis. *Journal of Experimental Biology*, *204*, 103–113.
- Kamil, A. C., & Jones, J. E. (1997). The seed-storing corvid Clark's nutcracker learns geometric relationships among landmarks. *Nature*, *390*, 276–279.
- Kamil, A. C., & Jones, J. E. (2000). Geometric rule learning by Clark's nutcrackers (*Nucifraga columbiana*). *Journal of Experimental Psychology-Animal Behavior Processes*, *26*, 439–453.

- Kitchin, R. M. (1994). Cognitive maps: What are they and why study them? *Journal of Environmental Psychology, 14*, 1–19.
- Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. L. (1998). Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *Journal of Neurophysiology, 80*, 425–446.
- Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage Publications.
- Leighton, J. P., & Dawson, M. R. W. (2001). A parallel distributed processing model of Wason's selection task. *Cognitive Systems Research, 2*, 207–231.
- McNaughton, B., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., Kudrimoti, H., Qin, Y., Skaggs, W. E., Suster, M., & Weaver, K. L. (1996). Deciphering the hippocampal polyglot: The hippocampus as a path integration system. *The Journal of Experimental Biology, 199*, 173–185.
- McNaughton, B. L., Barnes, C. A., & O'Keefe, J. (1983). The contributions of position, direction, and velocity to single unit-activity in the hippocampus of freely-moving rats. *Experimental Brain Research, 52*, 41–49.
- Muller, R. U., Bostock, E., Taube, J. S., & Kubie, J. L. (1994). On the directional firing properties of hippocampal place cells. *Journal of Neuroscience, 14*(12), 7235–7251.
- O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurones. *Nature, 381*, 425–428.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88*, 16–45.
- Redish, A. D. (1999). *Beyond the cognitive map*. Cambridge, MA: MIT Press.
- Redish, A. D., & Touretzky, D. S. (1999). Separating hippocampal maps. In N. Burgess, K. J. Jeffery, & J. O'Keefe (Eds.), *The hippocampal and parietal foundations of spatial cognition* (pp. 203–219). Oxford: Oxford University Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536.
- Sharp, P. E., Blair, H. T., & Cho, J. W. (2001). The anatomical and computational basis of the rat head-direction cell signal. *Trends in Neurosciences, 24*, 289–294.
- Shettleworth, S. J. (2002). Spatial behavior, food storing, and the modular mind. In M. Bekoff, C. Allen, & G. M. Burghardt (Eds.), *The cognitive animal* (pp. 123–128). Cambridge, MA: MIT Press.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*, 1–74.
- Spetch, M. L., Rust, T. B., Kamil, A. C., & Jones, J. E. (2003). Searching by rules: Pigeons' (*Columba livia*) landmark-based search according to constant

- bearing or constant distance. *Journal of Comparative Psychology*, 117, 123–132.
- Taube, J. S. (1998). Head direction cells and the neurophysiological basis for a sense of direction. *Progress in Neurobiology*, 55, 225–256.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55, 189–208.
- Touretzky, D. S., Wan, H. S., & Redish, A. D. (1994). Neural representation of space in rats and robots. In J. M. Zurada, R. J. Marks & C. J. Robinson (Eds.), *Computational Intelligence: Imitating Life*. New York: IEEE Press.
- von Frisch, K. (1966). *The dancing bees: An account of the life and senses of the honey bee* (2nd ed.). London: Methuen.
- von Frisch, K. (1967). *The dance language and orientation of bees*. Cambridge, MA: Belknap Press of Harvard University Press.
- Wehner, R., Bleuler, S., Nievergelt, C., & Shah, D. (1990). Bees navigate by using vectors and routes rather than maps. *Naturwissenschaften*, 77, 479–482.
- Wehner, R., & Menzel, R. (1990). Do insects have cognitive maps? *Annual Review of Neuroscience*, 13, 403–414.
- Worden, R. (1992). Navigation by fragment fitting: A theory of hippocampal function. *Hippocampus*, 2, 165–188.