The Handbook of Brain Theory and Neural Networks

EDITED BY
Michael A. Arbib

A Bradford Book
THE MIT PRESS
Cambridge, Massachusetts
London, England

Cognitive Development

James L. McClelland and Kim Plunkett

Introduction

Research in cognitive development seeks to understand how it is that the impressive cognitive capacities of the human mind can develop, and how this process is controlled. A central question has always been, Is the main source of control the innate endowment of the organism? Is it experience that arises from the interaction of the organism with its environment? Or is it an intricate interplay between genetic endowment and experience? These issues have been under consideration by philosophers and cognitive scientists for centuries. While no one today would deny the importance of either innate factors or experience, the exact nature of the mechanisms that exploit these factors, and the extent to which these mechanisms are preprogrammed with specific domain content, remains a source of intense debate within developmental psychology and cognitive neuroscience. In this article we consider how connectionist models may contribute to the evolution of thinking about these crucial matters.

Many connectionist models of cognitive development have been driven by two primary considerations:

1. *An interest in the role of environmental structure*. Developmental connectionists have tended to adopt a minimalist strategy by employing relatively general learning devices in specific problem domains in order to determine the extent to which the inherent structure of a domain constrains the construction of internal representations and thereby the developmental process.
2. *An attempt to specify the nature of the mechanisms involved in the developmental process*. In particular, connectionist models have tended to focus on the issue of whether discontinuities in behavioral development reflect the maturation of distinct information processors or nonlinear changes resulting from learning in a more homogeneous system.

Most connectionist models of cognitive development employ a general learning algorithm (such as backpropagation) that computes small changes to the connection strengths in a network so as to reduce the output error for any given input pattern. It is, therefore, appropriate to conceive of learning in these networks as a process of gradient descent on a multidimensional error landscape. Although the device that drives learning (the learning algorithm) usually only promotes gradual change in the weight matrix, the uneven surface of the error landscape can result in relatively sudden, dramatic qualitative shifts in network performance. Conversely, the error surface may be relatively flat at a particular configuration of the weight matrix, and the behavioral consequences of weight changes may be comparatively minor. Learning in networks can easily result in periods of stable behavior interrupted by sudden discontinuities, even though the basic mechanism for learning is one of small, continuous change.

The interpretation of change in a connectionist network as a process of movement through an even landscape is closely connected to Waddington's conception of development as epigenesis. Waddington (e.g., 1975) suggested that development can be viewed as a trajectory through a landscape in which stable states are achieved by an organism when occupying relatively homogeneous regions of the landscape, whereas change is observed during periods in which the trajectory crosses a downward sloping surface. The trajectory along the epigenetic landscape is determined by an interaction of the constitution of the organism itself and the environment in which it is required to survive. The same can be said of connectionist network models. What connectionist models add to Waddington's picture is a framework for constructing explicit models of the process of developmental change. Some of the properties of connectionist models, such as their use of graded parameters (connection weights) and gradient-based learning rules for adjusting these weights (such as backpropagation and many other connectionist learning rules) make them particularly well suited to the study of these issues. For fuller reviews of connectionist models and their importance for cognitive development, see Plunkett and Sinha (1992) and McClelland (1994).

It should be noted that connectionist models discussed in what follows apply to the development of implicit knowledge—knowledge that can govern behavior without itself being accessible to overt report. Knowledge in connections is implicit knowledge in just this sense. One key issue in cognitive development is the extent to which such implicit knowledge actually underlies performance in particular tasks. Clearly explicit rules and explicit reasoning strategies are sometimes used—a case in point will be suggested below. Developmental connectionists, however, have tended to stress that the behavior that others may have accounted for in terms of explicit rules might in fact be captured implicitly in connection weights.

Rethinking the Need for Innate Knowledge

We now consider the first general issue noted above, namely the extent to which cognitive processes must rely on innate domain knowledge. There can be no doubt that there is some initial structuring of the nervous system before birth that strongly influences what is experienced and the form this experience takes, and thus how this experience initiates changes in the structure of the system. However, there are several arguments that have been given by nativists that lead them to postulate innate knowledge of specific concepts or principles (see McClelland, in press, for fuller discussion). Here, we consider what we take to be the most crucial of these arguments, namely, the argument that a general-purpose learning mechanism that does not exploit domain-specific constraints is insufficient to account for the knowledge children acquire. This argument is based on the false assumption that the general-purpose learning mechanisms offered by connectionist learning rules adhere to the same principles as classical associationist models of learning: that learning occurs by contiguity, and application of what is learned to new cases depends on a similarity-based generalization process. Given this assumption, evidence that generalization depends on anything other than surface similarity is taken as evidence that there must be some domain-specific knowledge that is brought to bear. While there is considerable debate about the exact role that input or surface similarity might play and the sources of input that might count as providing data relevant to assessing input similarity, it nevertheless seems reasonably clear that the immediate perceptual characteristics of stimuli do not always determine the basis for correct generalization (Keil, 1987).

Where this argument goes wrong is in its initial assumption. Connectionist learning algorithms such as backpropagation are capable of gradually discovering task-appropriate representations through the learning process. Hinton's (1989) family tree model is one of the best examples of this: Before learning, this network treats all members of each of two "families" as approximately equally similar (with random initial variations). But after learning the kinship relations among these individuals through training with examples, the acquired internal representations use one dimension of the representational space to distinguish the two families, and use other dimensions to position the individuals within each family so that those individuals who play structurally analogous roles have similar representations. In other words, the similarity structure changes over time, gradually becoming appropriate to the domain through the course of the learning process. Another example that makes somewhat closer contact with the developmental literature is Rumelhart's model of learning the structure of the semantic domain of living things (Rumelhart and Todd, 1993). In this example, the representations of concepts gradually differentiate to capture a conceptual hierarchy that first distinguishes plants from animals and then later distinguishes among subclasses of these major categories (e.g., birds and fish). This process of conceptual differentiation is reminiscent of the developmental process of categorical differentiation discussed by Keil (1979).

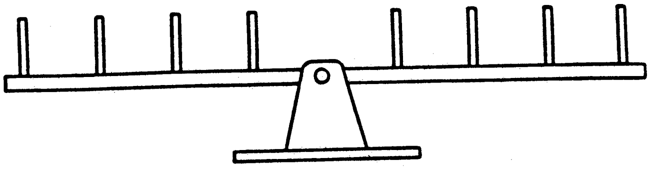

Figure 1. A balance scale of the type used first by Inhelder and Piaget and later by Siegler (1976, 1981). (Reprinted by permission from Siegler, 1976, fig. 1.)

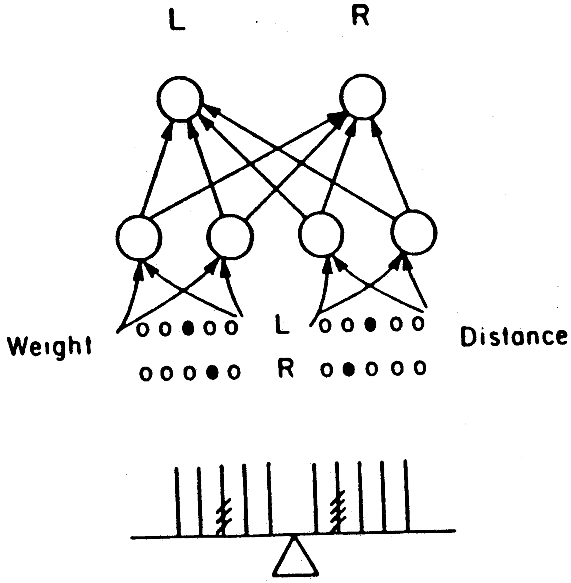

Figure 2. The connectionist network used to simulate acquisition of the balance scale task by McClelland (1989). Different input units code the number of weights placed on the Left and Right beams, and the positions at which they are placed. (Reprinted by permission from McClelland, 1989, fig. 2.7.)

Stages in Cognitive Development

We now consider the developmental process itself, in particular the issue of stage-like transitions in cognitive development. Stage theories of cognitive development characterize distinct stages of development in terms of qualitatively different underlying principles and operations. Transitions between stages are explained in terms of endogenous factors and/or a complex interaction between developing internal structures and the environmental niches which regulate the unfolding of those structures. In this section, we review a connectionist model of cognitive development whose behavior develops in a stage-like fashion. The properties of representation in such a model are interesting because input does not change over the course of training, and a single learning algorithm (backpropagation) with a fixed learning rate is used as the mechanism driving change throughout the simulation. An important finding from this simulation work is that a single mechanism can exhibit stage-like behavioral properties (i.e., there are accelerations and decelerations in the behavior and indeed in the connection weights in the system) despite the fact that the representational changes in the underlying mechanism are governed by a simple homogeneous process.

The domain of the model is the balance scale problem, introduced by Inhelder and Piaget (1958) and later studied extensively by Siegler (1976, 1981) and others (particularly Ferretti and Butterfield, 1986). Children are shown a balance scale with varying weights on either side and at varying distances from the fulcrum (Figure 1). They are asked to judge which side will go down when the scale is released. Siegler has shown that children pass through a series of stages in which their responses appear to be determined by a succession of procedures that make differential reference to the dimensions of weight and distance. Children below the age of 4 respond relatively haphazardly in the task. By 5, nearly all children are in the first stage, where they focus exclusively on the number of weights on each side in making a decision. In the second stage, they incorporate the distance dimension, but only under those conditions where the weights are equal. In the third stage, their responses are confused under those conditions where both weights and distance differ. Finally, some individuals eventually behave in accordance with a procedure that amounts to computing the torque exerted by the weights on each side of the scale by multiplication of weight times distance. Above the age of 5, a large proportion (approximately 93%) of children's responses to the balance scale task fitted into one of these four categories (see Siegler, 1976, 1981, for details).

The connectionist model that captured the developmental progression seen in the balance scale task was presented by McClelland (1989; Shultz, Mareschal, and Schmidt, 1994, have a different connectionist model of the same task). The input to the model (Figure 2) is divided into two channels—in this case, a channel that represents the weights of the two objects on either side of the fulcrum and a channel that represents the distance of two objects from the fulcrum. There are five possible weight values for each object and five possible distance positions for each object. Weight and distance values are represented for each object by a five-place input field, in which there is a single unit assigned to each integer value of weight or distance on each side of the scale. However, the weight and distance units are themselves unstructured—i.e., weight and distance values are arbitrarily assigned to single input units. The network is not told explicitly which units correspond to large weights or distances, nor which units represent the weight or position of the left or the right object. The network must discover these correspondences through experiencing outcomes of balance scale problems—i.e., by discovering connection weights that allow it to predict which side will go down for various combinations of object weights and distances.

The network is trained on random samples from the possible combinations of object weights and distances on the balance scale. Training proceeds gradually through connection weight changes made in response to a random sequence of such problems. On each trial, the network's output is compared to the correct output for that problem, and the discrepancy between the actual output and the desired output is used to generate an error signal which is used by a backpropagation learning algorithm to adjust the connection strengths in both channels of the network. The network is tested at regular intervals on both trained and novel weight/distance combinations.

Two factors are crucial to the performance of the network. First, the network must be structured so as to treat the dimensions of weight and distance as separate. Second, some differential treatment of weight and distance is necessary to reflect the fact that children rely on the weight cue earlier in development than the distance cue. In McClelland (1989), this differential treatment amounts to incorporating more examples involving weight variations than distance variations in the training set, on the assumption that children may have more experience with weight than with distance as a factor in determining balance. In McClelland (in press), it is shown that differential initial use of weight versus distance can arise if the weight on each side is a unary predicate (depending on the weight alone) while distance is a binary relation (a property arising from the relation between the weight and the fulcrum of the scale). Attention here focuses on the simpler case of more frequent exposure to variations in weight relative to distance. In both variants, the training sequence is stable over time. Discontinuity in development arises from the learning process, not from changes in the input.

At the beginning of training, the weights in the network are randomly set. Once training commences, the weight matrix gradually moves away from its random state, and after a time systematic behavioral patterns begin to appear. After an initial period of near-null output, there is a relatively rapid transition into conformity with the first of four rules developed by Siegler to characterize the stages of development of performance on the balance scale task. From this point on, the network's performance is impressive in that it can can be classified 85% of the time according to one of the first three of the four rules outlined by Siegler, and passes through the rules in the same sequence that characterizes children's development. The model does not reliably achieve the final stage of learning on the balance scale problem, which we take to be based on the use of an explicit multiplication strategy. (Siegler provides considerable evidence that performance in stage 4 is qualitatively different for crucial problems that rely on multiplication of weight times distance.)

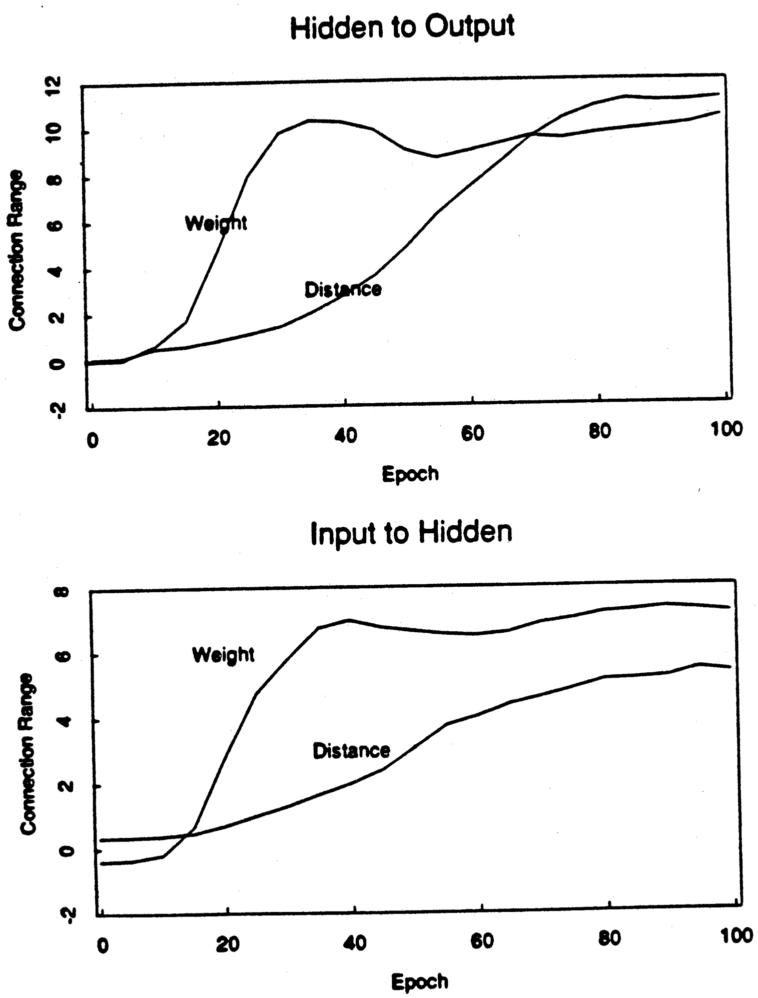

Figure 3. Sensitivity of the McClelland (1989) model to variations in weight and distance as cues to which side should go down on balance scale problems, as reflected by the input to hidden weights and by the hidden to output weights in the network (see Figure 2). The figure illustrates the periods of acceleration and deceleration seen in both pathways over the course of training. The index of sensitivity is based on absolute values of differences between connection weights from input to hidden units and from hidden to output units. For example, for sensitivity to object weight in the input to hidden connections, the graph shows the absolute value of the following difference: (the connection weight to the left-most hidden unit from the input unit coding for one weight placed on the left of the scale) minus (the connection weight to the same hidden unit from the input unit coding for five weights placed on the left of the scale). (Reprinted by permission from McClelland, 1989, fig. 2.12.)

Three aspects of the model commend the general approach it exemplifies for accounting for transitions in development in this and other domains:

1. The model shows periods of stasis followed by relatively abrupt transitions (Figure 3). Learning on the weight and then on the distance dimension is initially quite slow, then gradually accelerates until further changes fail to lead to further improvement in performance. The reason for the acceleration lies in the fact that the network must learn both how to encode the relevant dimension in the input to hidden weights, and how to use the results of this encoding to respond correctly (hidden to output weights). Changes at either level are effectively incoherent until the other level starts to become organized. This illustrates a crucial fact about development stressed extensively by Piaget: that the progress occasioned by a particular experience is sensitive to the existing state of knowledge, and that there must be a firm foundation before progress can be made (see Flavell, 1963, for discussion).

2. The model shows differential readiness to progress from one stage to the next at different points within the stage. Stage 1 typically lasts for quite a while in children, and the same thing is observed in the McClelland (1989) model. Both in the model and in children, exposure to a well-chosen set of problems leads to a transition to stage 2 or 3 if it occurs toward the end of stage 1, but either to no change or to regression to random performance if it occurs near the beginning. This differential readiness is a reflection of two factors in the model: first, the simple fact that weight changes accumulate, bringing the model closer to the point at which its behavior exhibits some sensitivity to distance; and second, the fact that the transition out of stage 1 corresponds to a point in development where weight changes in the distance pathway within the network accumulate rapidly. Very small changes build up during the early part of the time the network spends in a stage, leading to relatively abrupt transitions between stages.

3. Children's responses are not in fact exactly in line with the specific rules enumerated by Siegler, and the network captures many aspects of these discrepancies. Most interestingly, Siegler's rules are discrete, in the sense that they are not sensitive to different degrees of variation of either the weight or distance cue: The child either uses a cue (weight, distance) or not, according to Siegler's procedures. However, Ferretti and Butterfield (1986) and others have shown that children's behavior is strongly affected by the actual magnitude of the difference between the two sides on both the weight and the distance cues. McClelland (in press) shows that the model accounts quite well for these effects, whose relevance to the predictions of connectionist models was first pointed out by Shultz, Mareschal, and Schmidt (1994).

Finally, it may be noted that the model provides fairly clear links to the earlier work of Piaget (e.g., Piaget, 1952) in that it provides a simple but precise illustration of how a continuous developmental process of gradual adaptation can lead to periods of behavior that can be characterized as stage-like. The structural assumptions of the model require the assimilation of the input data to separate representations of weight and distance. The input assumptions of the model require the accommodation of the network's weights to sets of combinations of weights and distances which are repeatedly presented to the network. The interaction between structural and input assumptions can be seen as an embodiment of what Piaget would have called an *equilibration process*, that coordinates the representations of weight and distance in relation to the network's performance on the balance scale problem.

Discussion

The previous sections have shown that connectionist models provide powerful learning mechanisms that are capable of discovering appropriate internal representations. These mechanisms are more powerful than the learning mechanisms proposed by the classical associationists: in particular, they are not doomed to rely on raw input similarity as the basis for generalization. Because of their nonlinear, multilayer structure, connectionist models that learn are capable of exhibiting stage-like developmental progressions and other phenomena reminiscent of children's performance and of their developmental progress, as illustrated by the McClelland (1989) simulation of development in the balance scale task.

Connectionist models have also been applied to a number of other aspects of cognitive development. These include the development of "object permanence" (Munakata et al., 1994), the ability to order elements sequentially (Mareschal and Shultz, 1993), and the development of visually guided tracking and reaching (Mareschal and Plunkett, 1994). For a review of related connectionist-inspired models of language acquisition, see Language Acquisition. Taken together, these other models illustrate the potential breadth of application of the approach, and lead to the expectation that there will be more such models in the future.

The main issues that need to be addressed in further work are as follows:

1. How much prestructuring is actually needed to account for cognitive development? The McClelland (1989) model was prestructured to reflect the dimensional distinctions of weight and distance, and it will be important to consider whether such prestructuring is necessary and under what circumstances. It seems certain that we should view the brain as a developing network of networks, rather than a single, unstructured homogeneous system, and it seems likely that this network of networks receives some initial structuring prior to exposure to inputs from the external environment.
2. What is the relationship between the acquisition of implicit knowledge, captured by connectionist models, and explicit cognitive functions, including the ability to describe the basis for responses in tasks such as the balance scale task, and the ability to use this explicit knowledge as the basis for further processes? Many examples of one-trial learning would seem to rely on this form of explicit knowledge. (See Developmental Disorders for a discussion of this issue.)

It seems likely that these questions will motivate a great deal of future work, and that future developments in our understanding of learning processes in neural networks will shed considerable light on these issues.

Road Map: Connectionist Psychology
Related Reading: Developmental Disorders; Development and Regeneration of Eye-Brain Maps; Ocular Dominance and Orientation Columns

References

Ferretti, R. P., and Butterfield, E. C., 1986, Are children's rule-assessment classifications invariant across instances of problem types? *Child Dev.*, 57:1419–1428.

Flavell, J. H., 1963, *The Developmental Psychology of Jean Piaget*, Princeton, NJ: Van Nostrand. ◆

Hinton, G. E., 1989, Learning distributed representations of concepts, in *Parallel Distributed Processing: Implications for Psychology and Neurobiology* (R. G. M. Morris, Ed.), Oxford: Clarendon, pp. 46–61.
Inhelder, B., and Piaget, J., 1958, *The Growth of Logical Thinking from Childhood to Adolescence*, New York: Basic Books.
Keil, F. C., 1979, *Semantic and Conceptual Development: An Ontological Perspective*, Cambridge, MA: Harvard University Press.
Keil, F. C., 1987, Conceptual development and category structure, in *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization* (U. Neisser, Ed.), Cambridge, Eng.: Cambridge University Press, pp. 175–200.
McClelland, J. L., 1989, Parallel distributed processing: Implications for cognition and development, in *Parallel Distributed Processing: Implications for Psychology and Neurobiology* (R. G. M. Morris, Ed.), Oxford: Clarendon, pp. 9–45.
McClelland, J. L., 1994, The interaction of nature and nurture in development: A parallel distributed processing perspective, in *International Perspectives on Psychological Science*, vol. 1, *Leading Themes* (P. Bertelson, P. Eelen, and G. d'Ydewalle, Eds.), Hillsdale, NJ: Erlbaum, pp. 57–88. ◆
McClelland, J. L. (in press), A connectionist perspective on knowledge and development, in *Developing Cognitive Competence: New Approaches to Process Modeling* (G. Halford and T. Simon, Eds.), Hillsdale, NJ: Erlbaum.
Mareschal, D., and Plunkett, K., 1994, Developing object permanence: Connectionist insights, Paper presented at the Annual Conference of the Developmental Section of the British Psychological Society, September, Portsmouth, UK.
Mareschal, D., and Shultz, T. R., 1993, A connectionist model of the development of seriation, in *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum, pp. 676–681.
Munakata, Y., McClelland, J. L., Johnson, M. H., and Siegler, R. S., 1994, *Now You See It, Now You Don't: A Gradualistic Framework for Understanding Infants' Successes and Failures in Object Permanence Tasks*, Technical Report PDP.CNS.94, Pittsburgh: Carnegie Mellon University.
Piaget, J., 1952, *Origins of Intelligence in Children*, New York: International Universities Press. ◆
Plunkett, K., and Sinha, C. G., 1992, Connectionism and developmental theory, *Br. J. Dev. Psychol.*, 10:209–254. ◆
Rumelhart, D. E., and Todd, P. M., 1993, Learning and connectionist representations, in *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (D. E. Meyer and S. Kornblum, Eds.), Cambridge, MA: MIT Press, pp. 3–30. ◆
Shultz, T. R., Mareschal, D., and Schmidt, W. C., 1994, Modeling cognitive development on balance scale phenomena, *Machine Learn.*, 16:59–88.
Siegler, R. S., 1976, Three aspects of cognitive development, *Cog. Psychol.*, 8:481–520.
Siegler, R. S., 1981, Developmental sequences between and within concepts, *Monogr. Soc. Res. Child Dev.*, 46 (189).
Waddington, C. H., 1975, *The Evolution of an Evolutionist*, Edinburgh: Edinburgh University Press. ◆

Cognitive Maps

Nestor A. Schmajuk

Introduction

Cognitive maps store information about the relationships between contiguous temporal events or proximal spatial locations and combine this information to determine the relationships between remote temporal events or distant spatial locations. For example, during classical conditioning (see CONDITIONING), animals learn to predict what temporally contiguous events might follow other events. By linking several contiguous predictions, cognitive maps allow organisms to predict what temporally remote events might be expected. Similarly, during maze learning, animals learn to predict what spatially proximal locations are connected to other spatial locations. By linking several proximal predictions, cognitive maps allow organisms to predict what spatially remote locations are connected to other locations. This article outlines different formal theories and neural network models that have been proposed to describe both temporal and spatial cognitive mapping.

According to Tolman (1932), animals acquire an *expectancy* that the performance of response R1 in a situation S1 will be followed by a change to situation S2 (S1-R1-S2 expectancy). Tolman hypothesized that a large number of local expectancies can be combined, through inferences, into a *cognitive map*. Tolman proposed that place learning, latent learning, and detour learning illustrate the animals' capacity for reasoning by generating inferences. In place learning, animals learn to approach a given spatial location from multiple initial positions, independently of any specific set of responses. In latent learning, animals are exposed to a maze without being rewarded at the goal box. When a reward is later presented, animals demonstrate knowledge of the spatial arrangement of the maze, which remains "latent" until reward is introduced. Detour problems are maze problems that can be solved by integrating separately learned pieces of local knowledge into a global depiction of the environment.

When seeking reward in a maze, organisms compare the expectancies evoked by alternative paths. For Tolman, *vicarious trial-and-error behavior*, i.e., the active scanning of alternative pathways at choice points, reflects the animal's generation and comparison of different expectancies. At choice points, animals sample different stimuli before making a decision. For example, a rat often looks back and forth between alternative stimuli before approaching one or the other. According to Tolman's *stimulus-approach* view, organisms learn that a particular stimulus situation is appetitive, and therefore it is approached. Supporting this assumption, Mackintosh (1974:554) suggested that, in the presence of numerous intra-maze and extra-maze cues, animals typically learn to approach a set of stimuli associated with reward and to avoid a set of stimuli associated with punishment. However, in a totally uniform environment, animals learn to make the correct responses that lead to the goal.

Interestingly, Tolman (1932:177) suggested that the relations between initial and goal positions can be represented by a directed graph, and he called this graph a means-end field. Many years later, artificial intelligence theories described problem solving as the process of finding a path from an initial to a desired state through a directed graph (see PLANNING, CONNECTIONIST).

Place Learning

Place learning has been studied in the "water maze," requiring rats to escape from a pool filled with opaque water (Morris,