The Collins and Loftus model abandons the hierarchical structure used by Collins and Quillian in favor of a structure based on a person's experience. This means that the spacing between various concepts can differ for various people depending on their experience and knowledge about specific concepts.

In addition to proposing experientially based links between concepts, Collins and Loftus also proposed a number of additional modifications to the Collins and Quillian model to deal with problems like cognitive economy and the *pig/mammal* problem. The details of their proposed modifications aren't that important. What is important is that these modifications made it possible to explain just about any result of categorization experiments. Collins and Loftus describe their theory as "a fairly complicated theory with enough generality to apply to results from many different experimental paradigms" (1975, p. 427). Although you might think that being able to explain just about any result would be an advantage, this property of the model led some researchers to criticize it, as we will see in the next section.

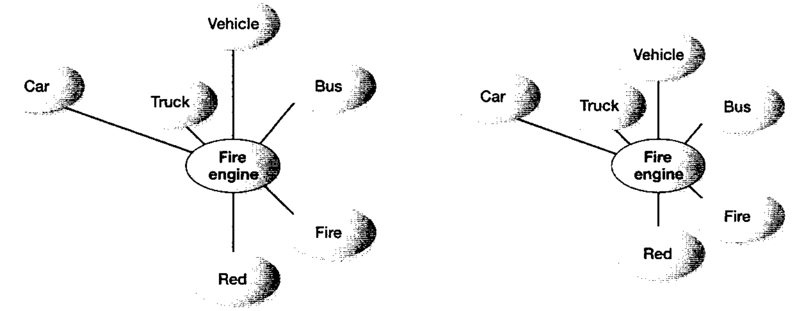
Assessment of Semantic Network Theories

Why would a model be criticized if it can explain just about any result? We can answer this question by considering the following properties of good psychological theories:

1. *Explanatory power.* The theory can explain why a particular result occurred by making a statement like "Behavior A occurred because . . ."
2. *Predictive power.* The theory can predict the results of a particular experiment by making a statement like "Under these circumstances, Behavior B will occur."
3. *Falsifiability.* The theory or part of the theory can potentially be shown to be wrong when a particular experimental result occurs. This means that it should be possible to design an experiment that can potentially yield results that would be predicted by the theory, and also that can potentially yield results that are *not* predicted by the theory.
4. *Generation of experiments.* Good theories usually stimulate a great deal of research to test the theory, to determine ways of improving the theory, to use new methods suggested by the theory, or study new questions raised by the theory.

When we evaluate the original Collins and Quillian theory against these criteria, we find that although it does explain and predict some results (see the data in Figure 8.15), there are many results it can't explain, such as the typicality effect and the longer reaction times for sentences like "A pig is a mammal." These failures to accurately explain and predict are what led Collins and Loftus to propose their theory.

But Collins and Loftus's theory has been criticized for being so flexible that it is difficult to falsify. We can understand why this is a problem by considering the networks in Figure 8.21, which show the node for *fire engine* and some of its links for two different people. The *fire engine* node would be more easily activated by related concepts for the network in (b) than in (a) because the links are shorter in (b). But the lengths



(a) Randy's links to fire engine (b) Sandra's links to fire engine

■ **Figure 8.21** The node for *fire engine* and some of the concepts to which it is linked for two different people: (a) longer links, and (b) shorter links.

of the links can be determined by a number of factors, including a person's past experience with fire engines or other types of vehicles. Unfortunately, there are no definite rules for determining these lengths—or, for that matter, for determining things like how long activation remains after it spreads, or how much total activation is needed to trigger a node. This means that by appropriately adjusting things like the length of the links and how long activation lasts, the model can "explain" many different results.

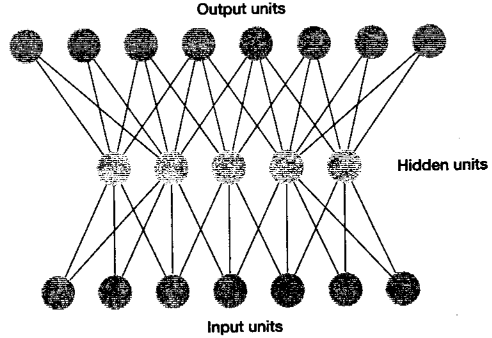
But if a theory can explain almost any result by adjusting various properties of the model, what has it really explained? That question is what led P. N. Johnson-Laird and coworkers (1984) to criticize semantic network theories and to conclude that these theories are "too powerful to be refuted by empirical evidence." This is a way of saying that it is difficult to falsify the theories. (See Anderson & Bower, 1973; Glass & Holyoak, 1975, for additional semantic network theories.)

Although research on semantic network theories was declining by the 1980s, network theories began a resurgence with the publication of two volumes titled *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* by James McClelland and David Rumelhart (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986). These books proposed a network model of mental functioning called **connectionism.**

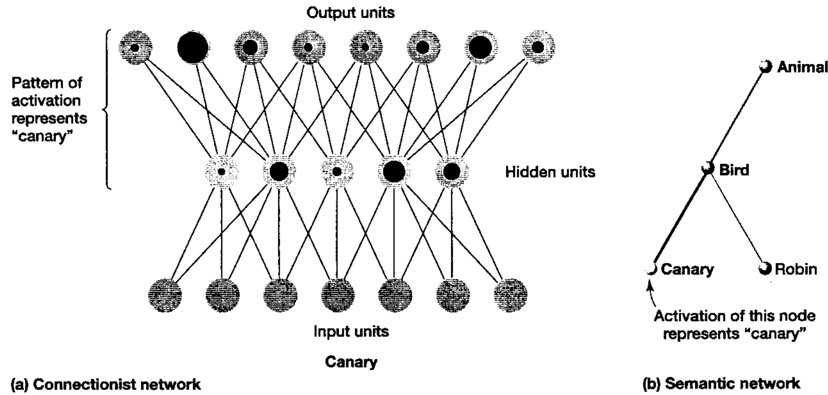
Representing Concepts in Networks: The Connectionist Approach

McClelland and Rumelhart proposed that concepts are represented in networks that contain nodes and links like semantic networks but that operate very differently from semantic networks. The idea of connectionist networks was partially inspired by the nervous system, in which neurons are connected to form neural networks (see Chapter 2). The basic characteristics of the connectionist networks, proposed by McClelland and Rumelhart, are:

- Connectionist networks consist of units, which are connected to form networks (McClelland & Rogers, 2003). McClelland describes these units as "neuron-like units" because they share properties with neurons. Like neurons, some units can be activated by stimuli from the environment, and some can be excited or inhibited by other units. Units are also connected with each other in circuits that resemble simple neural circuits (see Chapter 2). There are three types of units: **input units**, which are activated by stimulation from the environment; **hidden units**, which receive signals from the input units; and **output units**, which receive signals from hidden units (Figure 8.22).
- Knowledge is represented in connectionist networks by the distributed activity of many units (McClelland et al., 1995). Figure 8.23a shows how the concept *canary* might be represented by the pattern of activation in hidden and output units. What this means is that when a person thinks about what a canary is, the person's knowledge about the properties of canaries is represented in their mind by the pattern of activation in many units. Note that this is different from the situation in semantic networks, in which this knowledge about *canary* is represented by activity in individual nodes (Figure 8.23b).
- Because the processing in these networks, as in the nervous system, occurs in many parallel lines at the same time, and because the representation of concepts in these networks is distributed across many units, the connectionist approach is also called the **parallel distributed processing (PDP)** approach.
- Processing in connectionist networks is achieved by weights at each connection. Weights, which can be positive (analogous to excitation in neural circuits) or negative (analogous to inhibition), determine how strongly an incoming signal will activate the next unit.



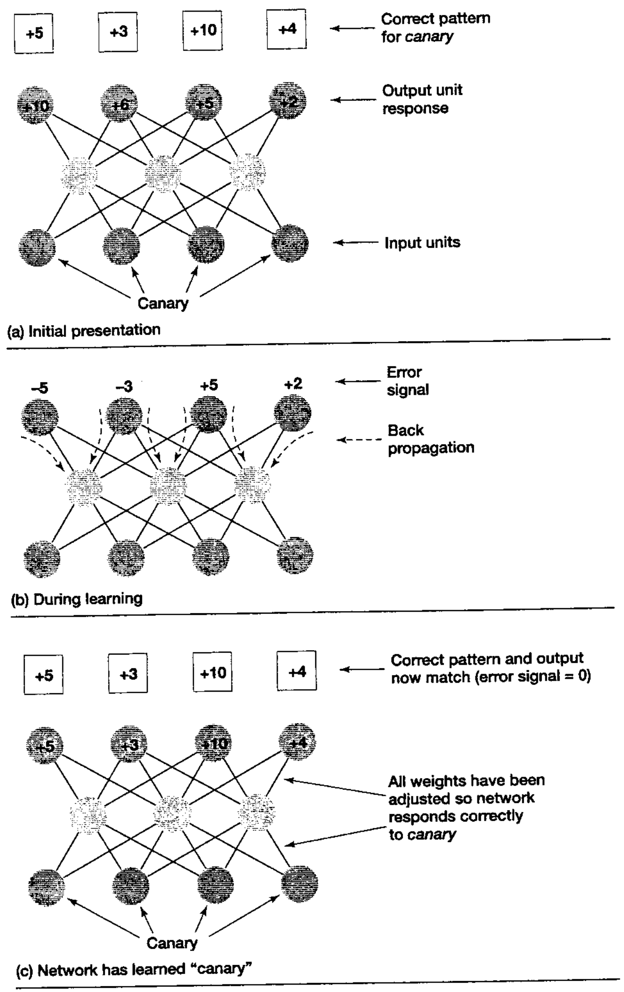
■ **Figure 8.22** A connectionist network showing input units, hidden units, and output units. Incoming stimuli activate the input units, and signals travel through network, activating the hidden and output units. Note that this is an extremely simplified version of a connectionist network. Networks used in research on connectionism contain many more units and more-complex connections between units.



■ **Figure 8.23** (a) How information is represented in a connectionist network by the distributed pattern of activity in a number of units. Activation is indicated by the size of the dots inside the units, with large dots indicating more activation. In this example, *canary* is represented by the pattern of activation of hidden and output units. (b) In a semantic network like the one in Figure 8.13c, *canary* is represented by activation of the *canary* node.

How does a pattern of activity in the hidden or output units become associated with a particular stimulus or concept? A number of different mechanisms have been proposed to explain how this occurs. We are going to focus on a mechanism called supervised learning, in which the network learns by a process that is analogous to the way a child gains knowledge about the world by making mistakes and being corrected. For example, a child learning language might point to a car and say *aumobile*. In response to this, the parent might provide the correct pronunciation—*auto-mobile*. The child usually continues to make mistakes, and may also mistakenly call a truck an *aumobile*. But eventually, with practice and continued guidance, the child learns how to say *automobile* and not to call trucks *automobiles*.

To see how a similar process occurs in a connectionist network, let's consider what happens when we present the input *canary* to the network in Figure 8.24. As you read this example, keep in mind that what we are describing is the response of a computer



■ **Figure 8.24** Learning in a connectionist network. See text for details.

that has been programmed to simulate a connectionist network, with its input, hidden, and output units.

To start the process, the experimenter specifies a concept, such as *canary*. Presenting *canary* excites the input units for canary and causes activation to flow through the network through the hidden units and then to the output units. Before any learning has occurred, the weights in the network are random, so the pattern of activity in the output units does not correspond to the correct pattern for *canary*. The correct pattern, which has been programmed into the computer by the experimenter, is indicated above the output units in Figure 8.24a. Just as a child uses the parents' correct pronunciation as a model for his or her next attempt at saying a word, the network uses the correct pattern as a guide for its next attempt at producing the pattern for canary.

The network uses this pattern to calculate an **error signal**, which is the difference between the actual activity of each output unit and the correct activity. The error signal for our example, which is indicated by the numbers above the output units in Figure 8.24b, is −5, −3, +5 and +2. This error signal provides information that the network can use to learn how to create the correct output pattern for *canary*. This learning occurs through a process called **back propagation**, in which the error signal is transmitted backward through the circuit. This back propagation of the error signal is symbolized by the dashed arrows in Figure 8.24b.

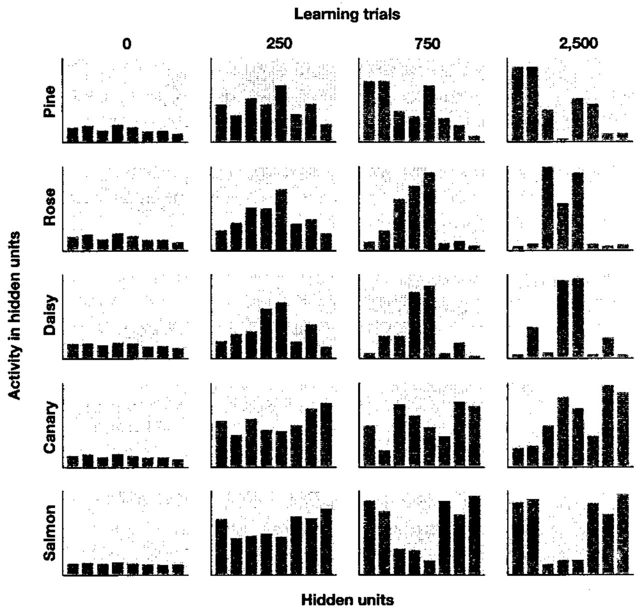
Information provided by the back propagated error signal indicates how the network's weights need to be adjusted so that the output signal will match the correct signal. From the error signal in our example, we can see that the strength of the inputs to the units on the left need to be decreased, and the strength of the inputs to two units on the right need to be increased. This is achieved by changing the weights of the connections between the units.

After the weights are changed, the input *canary* results in a new output pattern. This new pattern is closer to the correct pattern, but doesn't match perfectly. The process then repeats, with a new error signal being sent back through the network, and new weights being created, to bring the output pattern even closer to the correct pattern. Eventually, when the input *canary* results in an error signal of zero, the output is correct, and the learning process is completed (Figure 8.24c).

Although this process seems straightforward, with the network using the error signal on each trial to modify the weights and bring the output signal closer to the correct pattern, there is one complication: The network also needs to be able to respond correctly not only to *canary*, but to *robin* and *oak tree* and many other concepts. Thus, during the learning process, the networks' weights must be adjusted so it generates the correct pattern not only for *canary*, but also for *robin* and *oak tree* and other concepts as well.

One way to deal with this problem is to train the network on a large number of words or concepts at once, first presenting *canary*, then *robin*, then *oak tree*, and so on. Remember that the correct pattern will be different for each concept, with each one generating a different error signal. Because the network has to respond correctly to many different concepts, it is important to design the network's learning process so changing the weights to get a better response to *canary* doesn't result in a worse response to *oak* *tree*. This is achieved by changing the weights very slowly on each trial, so changing the weights in response to one concept causes little disruption of the weights for the other concepts that are being learned at the same time. Eventually, after thousands of trials, the weights in the network become adjusted so the network generates the correct output pattern for many different concepts.

Figure 8.25 shows how eight hidden units in a complex connectionist network respond during a learning process in which the network is presented with a number of different concepts, one after another (McClelland & Rogers, 2003). Each bar represents the activation in each of eight hidden units in response to different inputs. At the beginning of the process, activity is about the same in each unit (Learning trials = 0). But as learning progresses, with each concept being presented one after another and the weights being changed just slightly after each trial, the patterns become adjusted, so by Trial 250 the patterns for *salmon* and *canary* begin to look different, and by Trial 2,500 it is easy to tell the difference between the patterns for *salmon* and *canary* or between *canary*



■ **Figure 8.25** Patterns of responding in eight hidden units during learning in a connectionist network. See text for details. (Adapted from McClelland & Rogers, 2003.)

and *daisy*. Also note that the two flowers, *rose* and *daisy*, have similar but slightly different patterns.

There are two important things to remember about the connectionist approach: (1) It proposes a slow learning process that eventually creates a network capable of handling a wide range of inputs, and (2) information about each input is contained in the *distributed* pattern of activity across a number of units. Thus, just as the nervous system represents different faces by different patterns of activity in neurons (see Figure 2.13b), the connectionist network represents different concepts by different patterns of activity in its units. Some other properties of connectionist networks are as follows:

- *The system is not totally disrupted by damage.* Because information in the network is distributed across many units, damage to the system does not completely disrupt its operation. This property, in which disruption of performance occurs only gradually as parts of the system are damaged, is called **graceful degradation** and is similar to what often happens in actual cases of brain damage.
- *Learning can be generalized.* Because similar concepts have similar patterns, training a system to recognize the properties of one concept (such as *automobile*) also provides information about other, related concepts (such as *truck* and *vehicle*). This is similar to the way we actually learn about concepts because learning about automobiles enables us to predict properties of different types of automobiles we've never seen. This ability to generalize is the basis of intelligent behavior and the constructive nature of memory (see McClelland et al., 1995).
- *Successful computer models have been developed.* Computer models based on connectionist networks have been created that respond to being damaged in ways similar to the response that occurs in actual cases of brain damage in humans. Some researchers have suggested that studying the way networks respond to damage may suggest strategies for rehabilitation of human patients (Farah et al., 1993; Hinton & Shallice, 1991; Olson & Humphreys, 1997). In addition, connectionist networks have been developed that simulate normal cognitive functioning for processes such as language processing, memory, and cognitive development (Rogers & McClelland, 2004; Seidenberg & Zevin, 2006).

Although connectionist networks have a number of features that enable them to reproduce many aspects of concept formation, opinion regarding connectionist networks is divided. Some researchers believe that this approach holds great promise and are especially attracted to working on a system that shares some properties with the nervous system. Other researchers think that there are limits to what connectionist networks can explain, and feel that even if these networks may explain some aspects of how we store knowledge, the best way to explain how knowledge is represented in the mind is to combine connectionism with some of the other approaches to semantic memory that we discussed at the beginning of the chapter.

About the Author

E. Bruce Goldstein is a member of the cognitive psychology program in the Department of Psychology at the University of Pittsburgh and is Adjunct Professor of Psychology at the University of Arizona. He has received the Chancellor's Distinguished Teaching Award for his classroom teaching and textbook writing. He received his PhD in experimental psychology from Brown University and was a post-doctoral fellow in the Biology Department at Harvard University before joining the faculty at the University of Pittsburgh. Bruce has published papers on retinal and cortical physiology, visual attention, and the perception of pictures. He is the author of *Sensation & Perception* (7th edition, Wadsworth, 2007) and the editor of the *Blackwell Handbook of Perception* (Blackwell, 2001) and the forthcoming two-volume *Encyclopedia of Perception* (Sage). He teaches undergraduate courses in cognitive psychology and sensation and perception and a graduate course in the teaching of psychology.

E. Bruce Goldstein

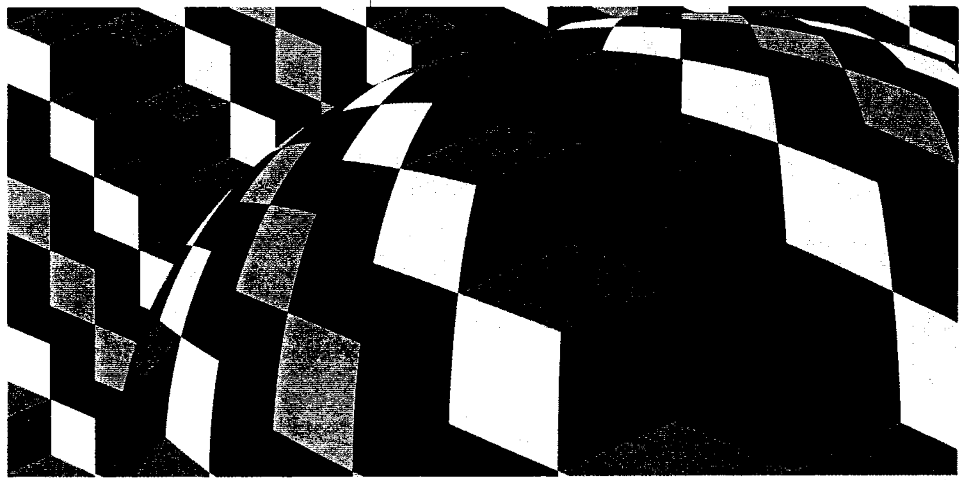
University of Pittsburgh

University of Arizona

SECOND EDITION

COGNITIVE PSYCHOLOGY

Connecting Mind, Research, and Everyday Experience



THOMSON

WADSWORTH

Australia • Brazil • Canada • Mexico • Singapore • Spain
United Kingdom • United States