OURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 71, 235–274 (1998)
RTICLE NO. CH982474

Neural Network Modeling of Developmental Effects in Discrimination Shifts

Sylvain Sirois and Thomas R. Shultz

McGill University

This paper presents neural network simulations of developmental phenomena in discrimination shifts. The discrimination shift literature is reviewed in order to identify the empirical regularities. Leading theoretical accounts of the development of shift learning are reviewed, and the lack of a thorough account is highlighted. Recent unsuccessful neural network simulations of shift learning are also reviewed. New simulations, using the cascade-correlation algorithm, show that networks can capture the regularities of the discrimination shift literature better than existing psychological theories. Manipulation of the amount of training that networks receive, which affects depth of learning, simulates developmental phenomena. It is suggested that human developmental differences in shift learning arise from spontaneous overtraining by older participants, an interpretation consistent with the overtraining literature. © 1998 Academic Press

Concept-shift tasks represent a useful benchmark for the developmental psychologist building computational models of human cognition. There are a substantial number of empirical data against which a model may be evaluated, as decades of research have identified robust phenomena (Kruschke, 1996). Moreover, such tasks involve both learning and cognitive development. For example, early research on concept-shift tasks and development has shown a relationship between shift performance and mental ability (Wolff, 1967). There are also qualitative distinctions between the performance of preschool children and that of older children and adults. The ability of artificial neural networks to capture the empirical regularities in this area of research may prove an important test of their adequacy as models of human cognition and its development.

This research was supported in part by a Natural Sciences and Engineering Research Council of Canada (NSERC) graduate fellowship to the first author and an NSERC operating grant to the second author. The authors thank Andy Baker, David Buckingham, Robbie Case, Yasser Hashmi, John K. Kruschke, Michael Rabinowitz, Yuriko Oshima-Takane, Yoshio Takane, and Sheldon H. White for their helpful suggestions. The authors also thank Charles J. Brainerd, John Rieser, and an anonymous reviewer for their comments on a previous version of the paper.

Correspondence concerning this article should be addressed to Sylvain Sirois, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montréal, Québec, Canada H3A 1B1. Electronic mail may be sent via Internet to sirois@psych.mcgill.ca.

0022-0965/98 $25.00

The validity of neural networks as models of human cognition was recently questioned from a developmental perspective (Raijmakers, 1996; Raijmakers, van Koten, & Molenaar, 1996). The authors carried out simulations of the balance scale and discrimination shifts (a subset of concept shift tasks) to highlight the inadequacies of feedforward neural networks as models of human learning. Because neural network models can be useful tools for the study of human learning and development (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996), we feel that the important issues raised by Raijmakers et al. with respect to concept-shift tasks deserve a response.

Our work focuses on three topics. The first is the hypothesis of age differences in shift learning. We review the psychological literature in order to identify the principal regularities and age differences in discrimination learning. The second topic concerns the traditional theoretical interpretations of age effects in discrimination learning and the lack of a comprehensive theoretical account. Finally, we evaluate the ability of artificial neural networks to simulate age effects in shift learning and the consequent implications to our understanding of human development.

DISCRIMINATION SHIFTS

Discrimination shift tasks represent an elementary form of concept-shift tasks (Wolff, 1967). Concept-shift tasks involve the categorization of stimuli according to attributes they exhibit on one or several dimensions. During the task, participants learn to identify which stimuli fall into which category based on these attributes. The target attributes are typically changed after an initial success criterion is reached. Stimuli may be presented individually (categorization) or simultaneously (discrimination). Typical responses in these tasks are sorting and verbal or motor identification. In discrimination shifts, the stimuli are presented in pairs and exhibit mutually exclusive attributes on all dimensions, and the criteria for correct responses change at some point during learning.

The motivation for restricting our focus to discrimination shifts is threefold. First, discrimination shifts are associated with a rich literature in which robust findings about human learning can be used as benchmarks by modelers (Kruschke, 1996). Second, the three leading theoretical accounts of concept-shift tasks, discussed later, originate from these discrimination tasks. Third, one purpose of this paper is to challenge the results obtained by Raijmakers et al. (1996), which were also restricted to discrimination shifts.

The discrimination shift paradigms we consider involve the pairwise discrimination of stimuli that exhibit attributes varying on three binary dimensions (e.g., shape, color, and position). In each pair of stimuli, members exhibit mutually exclusive attributes on all three dimensions. Figure 1 shows as an example the

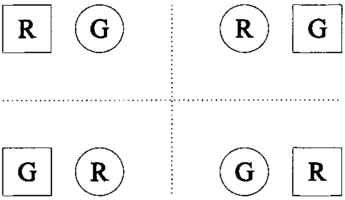


FIG. 1. An example of four stimulus pairs in a discrimination shift task. R and G denote red and green, respectively.

four pairs that exhaust the mutually exclusive combinations of shape, color, and position attributes.[1]

In such tasks, participants learn to identify the stimulus in each pair that exhibits the attribute targeted by the experimenter (e.g., the attribute *square* on the shape dimension). The irrelevant dimension may vary within all trials or may be held constant on a given trial but vary between trials. For example, if the irrelevant dimension is color, a red stimulus is always paired with a green stimulus in variable-within-trial paradigms, with shape varying as well. On the other hand, the red stimuli would be presented together and the green stimuli would also be presented together in variable-between-trials paradigms and only the relevant dimension (e.g., shape) would vary within a trial. Such a variant of discrimination shifts does not require mutual exclusivity of all attributes in a given pair.

Initial learning takes place by reward or lack thereof over repeated presentations of all pairs (see Fig. 2, second column). When participants reliably identify the target (typically, 8 out of 10 consecutive trials), they are considered to have successfully learned the initial discrimination. After criterion is achieved in this initial phase, several shifts of reward contingencies may be introduced.

One such shift is the reversal shift. In this condition, learning is shifted from the initial attribute to the other attribute of the same dimension (e.g., from *square* to *round*). Participants are therefore required to change their responses on all pairs. This is shown in Fig. 2 (second row).

Another possible shift is the nonreversal shift, shown in the third row of Fig. 2. In this condition, learning is shifted from the initial attribute to one from another, previously irrelevant, dimension (e.g., from *square* to *red*). With this shift, responses must be changed on only half of the pairs. Notice that half of the

[1] There is some confusion in the literature over whether these tasks involve 2^3 stimuli organized in 4 pairs that meet the mutual exclusivity constraint on attributes or rather 2^2 stimuli in two pairs, balanced for position. Gholson and Schuepfer (1979) suggested that some authors, neglecting position as a valid attribute of the stimuli, overlooked similarities in the behavior of young children and adults. For example, position bias in children has been mistakenly qualified as win-shift behavior because position was not included in analyses. We consider position as one dimension of variation in discrimination shifts, as did Tighe and Tighe (1972, 1978).

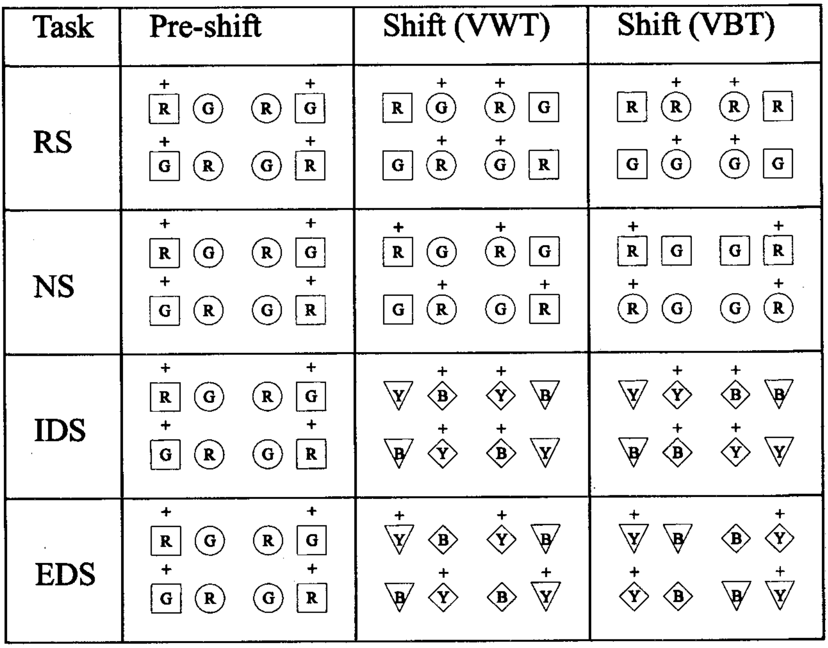


FIG. 2. Examples of four learning paradigms in discrimination shift: reversal (RS), nonreversal (NS), intradimensional (IDS), and extradimensional (EDS) shifts. Plus signs identify reward. For all four paradigms, initial training is with "square" as target, the irrelevant dimension (color) varying within trials. Columns 3 and 4 show changed reward contingencies for variable-within-trials (VWT) and variable-between-trials (VBT) shift conditions. R, G, B, Y denote red, green, blue, and yellow, respectively.

stimuli exhibiting the *square* attribute also exhibit the *red* attribute; responses to these pairs do not have to be changed.

There are two other paradigms that are similar to the reversal and nonreversal tasks just presented and can be labeled "total change" paradigms (Esposito, 1975). These tasks involve the introduction of new attributes of the dimensions at the onset of shift training. For example, red and green can be replaced by blue and yellow, square and round by triangle and diamond. An intradimensional shift involves a shift within the same dimension that was relevant during initial training (e.g., from *square* to *diamond*), a task similar to reversal shifts. This is presented in the fourth row of Fig. 2. An extradimensional shift, as the name implies, involves a shift to a previously irrelevant dimension (e.g., from *square* to *yellow*) and is thus similar to the nonreversal task (Fig. 2, bottom row).

The distinction between variable-within-trial and variable-between-trials paradigms is also shown in the third and fourth columns of Fig. 2. As can be seen, the irrelevant dimension varies within each pair for variable-within tasks, whereas it is held constant within a pair in variable-between tasks.

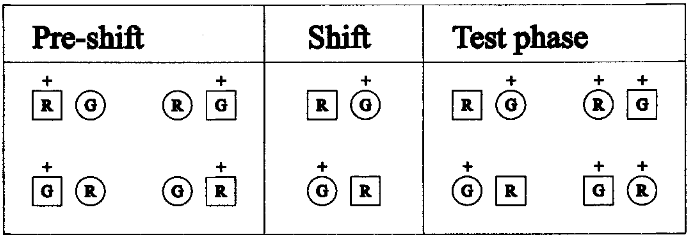


FIG. 3. An example of the optional shift task: initial learning (left), shift learning (middle), and test phase (right).

The final paradigm we focus on is the optional shift (Kendler, 1979, 1983; Kendler & Kendler, 1975; Wolff, 1967). This task has two learning phases (initial and shift) and one test phase. When criterion is reached in the initial phase, only half of the stimulus pairs are presented in the shift phase. These stimuli have shifted reward contingencies that are congruent with both a reversal and a nonreversal shift (Fig. 3, center). That is, whether the shift is within the original dimension (e.g., *square* to *round*) or to another dimension (e.g., *square* to *green*) cannot be assessed with only these pairs. After reaching criterion in the shift phase, participants are presented with all pairs in the test phase. For the stimulus pairs used in the shift phase, reward contingencies do not further change. But for the pairs previously used only in the initial phase, both stimuli are rewarded if selected (Fig. 3, extreme right). The purpose of this task is to evaluate whether or not the behavior of participants on the test pairs follows the pattern of a reversal shift with respect to initial training (e.g., in Fig. 3, extreme right, choosing the "red and round" stimuli and not the "green and square" ones). If a participant does so consistently (usually, 8 or more times out of 10), he or she is labeled "reverser." Otherwise, he or she is labeled "nonreverser."

PSYCHOLOGICAL REGULARITIES

Most normal human adults, and children above the age of 10, execute a reversal shift faster than a nonreversal shift (Esposito, 1975; Kendler, 1983, 1995; Kendler & Kendler, 1975; Kruschke, 1996; Wolff, 1967). When they are presented with a nonreversal shift, their performance on the unchanged pairs, initially correct, nevertheless drops substantially before it improves for the criterion run (Tighe & Tighe, 1978). Finally, about 87% of adults and older children exhibit reversal behavior on the test pairs of the optional shift and are labeled "reversers" (Kendler, 1983).

These findings contrast with those observed in preschoolers. Namely, about 80% of preschool children do not show reversal behavior on the test problems of the optional shift and are labeled "nonreversers" (Kendler, 1983, 1995). When they perform a nonreversal shift, their performance on the unchanged pairs remains high throughout (Tighe & Tighe, 1978). Finally, the literature does not suggest an advantage for either type of shift in comparisons of reversal and nonreversal shifts in children. Some studies report reversal superiority, some report no difference, and some others report nonreversal superiority (Esposito, 1975; Wolff, 1967). The stimuli used may play a large role in this, as children prefer some dimensions or attributes over others, and thus learn differently depending on whether their preferred dimension or attribute is the target or not (Esposito, 1975). Given the scholarly journals' emphasis on significant results (Lips, 1993), the safest conclusion is that there is no overall difference in preschoolers' performance on reversal and nonreversal tasks and that those reported are experiment specific (Esposito, 1975; Wolff, 1967). This is related to the controversy over whether there are regularities in younger children's shift behavior (Esposito, 1975). Variability of ease of reversal versus nonreversal shifts within ages as well as variability within individual children seems to be the rule (Cole, 1973, 1976). Even with this variability, reversal shifts become easier than nonreversal shifts between the ages of 4 and 10 (Esposito, 1975; Wolff, 1967). It is worth noting that when preschoolers are trained for several trials beyond the usual success criterion (ranging from 10 to 100 additional trials), they perform a reversal shift faster than a nonreversal shift as older children and adults do (Wolff, 1967). This is known as the overtraining effect.

The suggestion that there is no difference between reversal and nonreversal shifts in younger children is in contradiction with the pervasive assumption that preschool children perform a nonreversal shift faster than a reversal shift (Kendler, 1979, 1983; Kendler & Kendler, 1975; Raijmakers et al., 1996). However, for reversal and nonreversal shift tasks in which the same stimulus pairs are used in all learning phases and participants are not explicitly advised about the introduction of the shift, the literature does not support nonreversal superiority over reversal in kindergarten children.

The initial finding of nonreversal superiority by the Kendlers (Kendler & Kendler, 1959; Kendler, Kendler, & Wells, 1960; Wolff, 1967) was obtained with a task in which the stimuli were paired, at the onset of the nonreversal shift, according to their attribute on the originally relevant dimension, a variable-between-trials paradigm (e.g., both "large" figures together, both "small" figures together). This situation not only makes the onset of the shift obvious to participants, but also gives a reliable clue about the dimension that is now relevant. The reversal shift task, on the other hand, used a variable-within-trial paradigm (i.e., relevant and irrelevant dimensions varied within trials). Such confounds do not warrant equitable comparison of reversal versus nonreversal shifts. In another experiment, they introduced new stimuli in the nonreversal condition that kept the previously relevant dimension constant and introduced a new dimension as relevant (Kendler et al., 1960). Again, younger children found the nonreversal shift easier than the reversal shift.[2] But this experiment has the

[2] With such a task, college students also find the nonreversal shift easier than the reversal shift (Wolff, 1967).

same two limitations as the first one. The Kendlers were unable to obtain nonreversal superiority over reversal with the standard task and thus concentrated their developmental research on the optional shift task (Kendler & Kendler, 1975). In optional shifts, reversers perform consistently with a reversal shift in the test phase whereas nonreversers do not. Data from optional shift tasks have been used to support the assumption of nonreversal superiority in young children, essentially because two limitations have been overlooked. What this task really evaluates is whether participants generalize a shift to test items. When they do generalize, the shift is consistent with a reversal shift. Beyond this information, optional shifts warrant no conclusion about the ease of reversal shifts over nonreversal shifts.[3] Another limitation is that different criteria are used to classify reversers and nonreversers. Although reversers must exhibit behavior consistent with a reversal shift (≥80% of the test trials), nonreversers are simply not reversers. They are not required to consistently show nonreversal-compatible behavior. This should further caution researchers about the relevance of optional shift data for reversal and nonreversal shift comparisons.

THEORETICAL INTERPRETATIONS

The three major theoretical accounts of human shift learning are those of the Kendlers (Kendler, 1979, 1983; Kendler & Kendler, 1962, 1969, 1975), Tighe and Tighe (1966a, 1966b, 1972, 1978), and Zeaman and House (1963, 1974, 1984). Both the Kendlers and Zeaman and House stressed that mediated processing underlies adult performance. Whereas Zeaman and House (1974) suggested that mediated processing is also involved in preschool children, the Kendlers argued that young children form simple associations between stimuli and overt behavior (Kendler, 1983; Kendler & Kendler, 1975). Tighe and Tighe (1966a), on the other hand, offered a perceptual interpretation without mediation.

The Kendlers suggested that preschoolers, like other animals, learn discriminations by means of associative processes (Kendler, 1983, 1995; Kendler & Kendler, 1969, 1975). They argued that young children do not specifically encode the relevant attributes of the stimuli, but rather associate the compound properties of the stimuli with a response. But as children grow older, developmental changes allow them to use categorical responses to mediate the processing of stimuli into overt responses. These categories represent the relevant dimensions involved in the problem. By reinforcing the responses related to the appropriate category, adults find the reversal shift easier because only the link between the category and an overt response needs to change. In a nonreversal shift, the responses from the initial category need to be extinguished and responses to another category have to be trained. This would require more training than in the reversal task. These mediating categories facilitate rapid

[3] An integrated account of discrimination shifts, though, should obviously account for optional shifts in a manner consistent with its account of reversal and nonreversal shifts.

reversal shifts and also explain the reversal behavior of adults in the optional shift task.

Zeaman and House (1974, 1984) argued, on the other hand, that the mediational processes between stimuli presentation and overt behavior are of an attentional nature. When learning a discrimination, participants attend to one dimension and act upon it. Following a reward, links between stimuli and attentional responses, as well as those between attentional responses and overt responses, are strengthened. Different parameters modulate these two types of learning. When participants reach criterion, they have learned to attend to the relevant dimension, as well as to make the appropriate response within that dimension. Following the same logic as that found in the Kendlers' account, shifts within the same dimension are easier than those to a previously irrelevant dimension. What is novel in this account is that preschoolers are also believed to perform according to this dual-level processing (Zeaman & House, 1963, 1974). This implies that kindergarten children would also find a shift within the initial dimension easier than a shift to another dimension.

This discrepancy between the two interpretations stems from the fact that Zeaman and House introduce new stimuli at the onset of the shift (Tighe & Tighe, 1978; Wolff, 1967; Zeaman and House, 1963). An intradimensional shift is similar to a reversal shift, because training is shifted to an attribute within the same initial dimension; and an extradimensional shift involves a shift to a previously irrelevant dimension, as is the case for the nonreversal shift (Fig. 2). Young children learn something about the initially relevant dimension that can be transferred to new stimuli. Specifically, younger children execute an intradimensional shift faster than an extradimensional shift, a fact that the Kendlers' theory cannot account for (Esposito, 1975). The categories in the Kendlers' account represent the discrete attributes used in the experiment (e.g., "large"), whereas Zeaman and House argue that participants attend to the perceptual dimensions (e.g., "size") rather than their values. On the other hand, the Zeaman and House model suggests that young children perform a reversal shift faster than a nonreversal shift, a fact that is not supported. Tighe and Tighe (1978) argued that the two theories arose from experiments that stressed different processes. Rather than being contradictory, they should be considered complementary. However, these two theories have yet to be integrated.

Tighe and Tighe (1966a, 1978) offered a third account of discrimination shift, in response to the limitations of mediation theories, based on differentiation theory (Gibson & Gibson, 1955). Rather than emphasizing internal representations as the core feature of mature performance, Tighe and Tighe suggested that Gibsonian perceptual differentiation would better account for human shift learning because of the emphasis on stimulus properties (Tighe & Tighe, 1966a). Discrimination learning thus involved the identification of invariants in a sequence of stimuli, consistent with the reward contingencies. Adults perform the reversal and intradimensional shifts easier than nonreversal and extradimensional shifts, respectively, because of learned differentiated associations to the relevant and irrelevant dimensions. Poor differentiation in preschoolers, though, would account for ease of nonreversal shifts over reversal shifts. The relevant dimension would not be isolated from the irrelevant one. Such compound encoding of stimuli makes the nonreversal shift easier, because only half of the responses need to be changed. Development would involve a change from object to dimensional control. Children would still perform the intradimensional shift easier than the extradimensional shift, because the poorly isolated invariants would not be confounded with the initial stimuli at the onset of the intradimensional shift. In total change paradigms, initial compound responses to stimuli cannot be used in the shift phase of the experiment, and thus interference is removed from a minimal differentiation that was acquired during initial training.

Tighe and Tighe's (1966a) theory marks a departure from the other two. While it attempted to answer the distinct findings highlighted by the mediational approaches, it also removed mediation from the process.[4] Overt behavior is a direct function of differentiation. That is, association of stimuli and responses is based on perceptual properties, without mediating responses. Yet there are two important limitations to this theory. The first is that, as noted earlier, reversal and nonreversal shifts should be as easy (or as difficult) for preschoolers. There is no empirical support for Tighe and Tighe's (1966a) suggestion of nonreversal superiority in younger children. A second limitation, which poses a problem for mediational accounts as well, is its incapacity to account for dimensionless shift learning.

Dimensionless shift experiments represent another key element in the lack of a unified theory of discriminative learning. Adult research initiated by Bogartz (1965), as well as developmental studies by Goulet and Williams (1970), has shown that the presence of categories or dimensions in the stimuli was not necessary for reversal behavior in adults, nor its development in children. When learning the meaningless categorization of stimuli into two arbitrary categories (e.g., pictures of unrelated objects) or learning to discriminate meaningless stimuli (e.g., nonsense syllables or shapes), adults execute a full reversal easier than a half reversal (Bogartz, 1965). That is, changing all categorical responses, as in a reversal shift, is easier than changing half of the responses, as in a nonreversal shift. The development toward easier reversal over nonreversal was also observed in school-age children using meaningless stimuli or categories and full versus half reversal tasks (Goulet & Williams, 1970). Because the stimuli

[4] Tighe and Tighe's (1966a) model involves a transformation of the stimuli into the overt response, which in a way can be construed as a form of generic mediation (as one reviewer observed). The stimuli properties are combined and transformed into a response, a process we can refer to as "simple mediation." In this paper though, the term mediation is used in the usual sense found in the discriminative learning literature: a transformation of the stimuli into a different representational format, which is then transformed into an overt response. Mediation is meant to involve an intermediate response. In the remainder of this paper, mediation thus implies two levels of transformations and an intermediate response, which is not found in Tighe and Tighe (1966a).

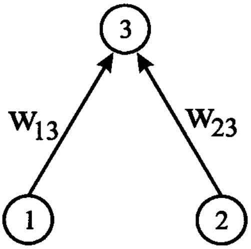


FIG. 4. An example of connected units in a network. Units 1 and 2 send their activations to unit 3 through their weighted connection w_{ij}.

cannot be categorized on the basis of perceptually differentiable features, Tighe and Tighe's theory cannot account for these data. Neither can the two mediational theories, because the categories are not based on conceptual dimensions. Because of this general shortcoming, as well as each theory's own limitations, the fact remains that there is no successful general account of shift learning to this day.

CONNECTIONIST MODELING OF DISCRIMINATION SHIFTS

Artificial neural networks consist of simple processing units that transmit their activation to the other units in the network through weighted connections. Figure 4 presents two such units that are connected to a third unit. The activation of a unit is modified by the net input that it receives from other units. Activations of sending units are multiplied by the corresponding connection weights and these products are summed. The net input to the receiving unit, at a given time, is computed by

$$\text{net}_{a_i} = \sum_j w_{ij} a_j ,$$

where i is the index of the sending unit, a is the activation of the sending unit, and w is the connection value between the sending and receiving units. The net input of a unit can be understood as the total amount of stimulation or inhibition, summed across all sending units, that the unit receives at a given time. The output value of this receiving unit can be equal to the net input in the case of linear activation units (an identity function). More often, the net input is squashed through a nonlinear activation function. This is typically the asigmoid function, such as

$$a_i = \frac{1}{1 + e^{-\text{net}_{a_i}}}$$

which has an S shape, as presented in Fig. 5. Such nonlinear functions are typically used for hidden and output units in feedforward networks.

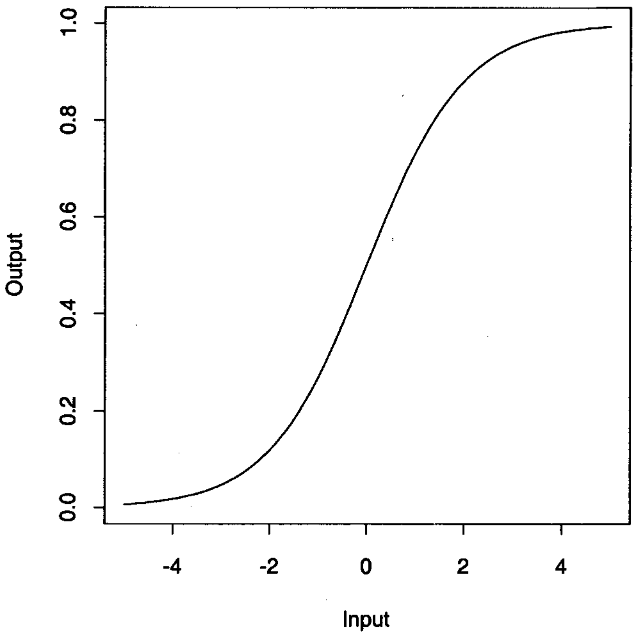


FIG. 5. Plot of y as an asigmoid function of x.

Feedforward networks are a special class of neural networks for which activation flows in only one direction. A layer of input units receives activation from the environment. In the simplest case, these activations are directly propagated to the output units (Fig. 6, left). Such networks are called perceptrons (i.e., Rosenblatt, 1958), and they can learn functions that are linear combinations of the input values (i.e., linearly separable problems). Perceptrons cannot learn nonlinear functions, but multilayer feedforward networks can. Multilayer networks have at least one layer of hidden units (Fig. 6, center). These units are labeled hidden because they do not receive direct activation from the environment nor produce external output. Activations from these hidden units are sent either to another layer of hidden units or to the output units. Finally, multilayer feedforward networks can include direct connections between input and output units as well as hidden layers (Fig. 6, right).

Neural network simulations of discrimination shift learning that would utilize mediation would presumably employ hidden units. The activation patterns on hidden units would mediate the relation between the input units (stimuli) and the output units (responses).

Feedforward networks (perceptron and multilayered) can learn to reproduce desired output for specific input patterns by changing the connection weight values between units. Typically, the learning rule performs error reduction,

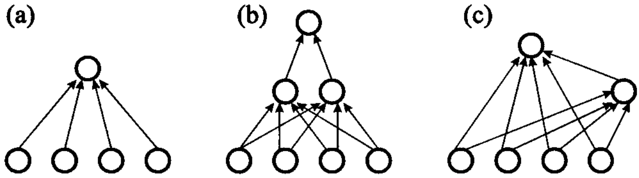


FIG. 6. Three different types of feedforward architectures: a perceptron (a), a multilayered network (b), and a multilayered network with cross-connections (c).

where error is the discrepancy between the desired output and the actual output produced by the network.

Because feedforward connectionist networks can learn, discrimination shifts provide a good benchmark for their adequacy as models of human cognition. A worthy neural network model of these tasks could be useful in formulating a general account of human shift learning. Surprisingly, there are few published neural network studies of shift learning, and within these there is no successful account of discrimination shifts. Kruschke (1996), for example, modeled compound categorization shifts with a connectionist model that qualitatively fits the psychological data. These tasks are different from discrimination shifts, though. Stimuli are presented individually, and categories involve relationships between stimulus attributes.

Raijmakers et al. (1996), on the other hand, suggested that feedforward neural networks cannot capture adult behavior on discrimination shifts and are therefore inadequate models of human learning. Because feedforward networks are bottom-up associative systems, the conjecture that the networks cannot capture the top-down conceptual behavior hypothesized in human adult discrimination is a legitimate working hypothesis. Raijmakers et al. were concerned with whether neural networks would behave like human adults or like preschoolers in a learning task. To assess this, they submitted neural networks to reversal, nonreversal, and optional shift tasks.

Their feedforward networks used the backpropagation error correction algorithm and consisted of three layers. For all networks, there were eight input units. The first two coded color of the left stimulus, the next two shape of the left stimulus, the following two color of the right stimulus, and the last two shape of the right stimulus. These units received binary input. The hidden layer consisted of either two (8–2–2 networks) or four hidden units (8–4–2 networks). Some networks had connections to hidden units from all input units (unconstrained networks), whereas the remaining networks had dimension-specific connections from input to hidden (constrained networks). These constrained hidden units received input from only one dimension. For all topologies, hidden units were connected to two output units. Learning involved turning on the appropriate output unit for each input pattern.

Most of their networks (8–4–2 constrained, 8–4–2 unconstrained, and 8–2–2 unconstrained) learned a nonreversal shift faster than a reversal shift. There was no difference between reversal and nonreversal shifts for the 8–2–2 constrained networks, but a trial-by-trial analysis showed that they performed like preschool children during shift learning (i.e., high performance on unchanged nonreversal pairs, and equally low performance on reversal and changed nonreversal pairs). In the optional shift task, all types of networks responded in a nonreversal manner, following the preschool data. Overtraining did not help the networks perform as adults. Raijmakers et al. (1996) concluded from these experiments that "the learning behaviour of feedforward PDP networks with error backpropagation in all tested configurations appears to be better described as making direct connections between stimulus and response rather than making connections by way of mediating concepts" (p. 128).

The first question we raise is whether their network topologies were justified. Although the authors did not explain why they chose three-layer networks (we assume that it was to model mediated processing), it can be argued that their use of hidden units was unnecessary. In their simulations, the desired output is a linear function of the input (i.e., the problems were linearly separable). A two-layer perceptron that includes only input and output units can therefore solve their tasks. Forcing a network with nonlinear hidden units between input and output to learn linear functions makes learning more difficult than it ought to be.[5] This is because net inputs to hidden units are submitted to unnecessary nonlinear transformations and because the error signal is diffused through a more complex structure. It is at best awkward, and at worst impossible, for nonlinear transformations to implement linear functions. The topology they elected to use thus had an impact on their results.

A second concern is that their networks learned the nonreversal shift faster than the reversal shift. Although some researchers believe this to be true of young children, the psychological literature does not support this, as noted earlier. Even though their networks did not behave like human adults, we believe that they did not behave like preschoolers either, because as noted preschoolers learn the two shifts equally fast.

We suggest that these limitations do not permit the conclusion that artificial neural networks are inadequate models of human learning. But there is still the challenge to successfully model human performance in discrimination shifts and provide a general theoretical account of performance on these tasks. The next section presents our simulations of discrimination shifts.

A CASCADE-CORRELATION MODEL

Cascade-correlation (Fahlman & Lebierre, 1990) is a feedforward algorithm that alters the topology of the network as it learns. At the onset of training,

[5] The question of what topology to use is problematic for static networks. To model human performance, the topology is often selected by trial and error. In this case, perhaps it was assumed that mediational accounts were correct; thus only networks with hidden layers were tried.

cascade-correlation networks have a minimal perceptron-like architecture (i.e., there are no hidden units). As the networks learn, they recruit the hidden units required to solve the problem. Networks learn by changing their weight values (or connection strengths) in order to minimize the discrepancy between their current output and a desired output (i.e., feedback about their performance on a given trial). When learning stagnates, because the networks lack the sufficient computational power, the algorithm installs hidden units that are trained to track residual error. For a discussion of the generative aspects of cascade-correlation, as well as for the mathematics of the learning rule, the reader is referred to Mareschal and Shultz (1996).

From a developmental perspective, the structural plasticity of cascade-correlation networks offers some advantages over static architectures (Mareschal & Shultz, 1996). Networks can adapt to the specific problems they attempt to learn. Cascade-correlation has been used to successfully model a wide range of developmental phenomena (Mareschal & Shultz, 1996), including seriation (Mareschal & Shultz, 1993), the acquisition of velocity, time, and distance concepts (Buckingham & Shultz, 1994), conservation (Shultz, 1998), the acquisition of personal pronouns (Shultz, Buckingham, & Oshima-Takane, 1994), and the balance-scale (Shultz et al., 1994).

This is the first developmental extension of this algorithm to a learning task (i.e., where learning is assessed during the experiment and not over development). Although we could have used static backpropagation networks for these simulations, the plasticity of cascade-correlation and its previous successes at modeling developmental phenomena make it a candidate model for a yet to be formulated theory of cognitive development (Mareschal & Shultz, 1996; Shultz & Mareschal, 1997). Moreover, the generative property of cascade-correlation gives researchers the advantage of not having to explore or design topologies by hand as Raijmakers and her colleagues had to do.

Because discrimination shifts are linearly separable problems from a modeling perspective, the architectures of adult cascade-correlation networks are identical to those of child networks. Namely, both contain only input and output layers of units because cascade-correlation does not install hidden units that are not required. We thus need to identify parameters that could distinguish adult from preschool networks.

Psychological literature has highlighted differences between kindergarten children and older children in the quality of their learning (Case, 1978; Siegler, 1978). Specifically, older children learn more accurately than younger children with equivalent training and show better generalization (Case, Kurland, & Goldberg, 1982). In general, older children and adults are considered to make finely tuned distinctions that younger children do not.

One mechanism that is believed to be involved in such change is rehearsal (Craik & Lockhart, 1972; Hagen, Jongeward, & Kail, 1979; Siegler, 1998). Between the ages of 5 and 10, there is substantial improvement in memory task performance related to increased spontaneous rehearsal (Flavell, Beach, & Chinsky, 1966). Under certain conditions, experimentally induced rehearsal in preschool children improves their performance (Brainerd, Olney, & Reyna, 1993; Hagen, Hargrove, & Ross, 1973; Hagen et al., 1979). And learning ability, assessed by tasks that tap short-term memory resources, shows an important improvement between the ages of 4 and 11, after which there is no further significant gain (Inglis, Ankus, & Sykes, 1968).

Ornstein and his colleagues identified conditions under which induced rehearsal may impair recall; namely, younger children rehearse individual items in isolation (passive rehearsal), which reduces recall for lists of unrelated items (Naus, Ornstein, & Aivano, 1977; Naus, Ornstein, & Kreshtool, 1977; Ornstein, Naus, & Liberty, 1975; Ornstein, Naus, & Miller, 1977). On the other hand, simultaneous rehearsal of several items (active rehearsal), as observed with older children, is associated with better recall, especially if the items are semantically grouped in the rehearsal set (Ornstein et al., 1975, 1977). These data suggest that rehearsal content, rather than frequency, is associated with better recall (Ornstein et al., 1975). Yet frequency or amount of rehearsal increases with age. We thus suggest that rehearsal may be related to the ontogeny of shift learning. Rehearsal would effectively result in more trials.

There is a parallel to induced rehearsal in the discrimination shift literature called overtraining (Kendler, 1983; Raijmakers, 1996; Wolff, 1967). This technique involves extra training trials beyond the usual success criterion. In young children, this additional training has little effect on their nonreversal shift performance but it facilitates the learning of a reversal shift (Wolff, 1967). Overtraining also increases the likelihood of reversal behavior in children performing an optional shift (Kendler, 1983; Tighe & Tighe, 1966b; Wolff, 1967).

Both overtraining and induced rehearsal are external constraints that result in more processing for participants. The outcomes are similar: in both cases, performance is affected through extended exposure to materials. A possibility, then, is that older children and adults learning discrimination shifts spontaneously submit themselves to overtraining through a form of iterative processing. What would be processed iteratively is the relationship between stimulus properties, responses, and feedback.

This idea has already received support in Levine's hypothesis-testing theory (Levine, 1966, 1975). He used a four-dimensional discriminative learning task with several blank trials (i.e., no feedback). Levine analyzed the responses of participants during the blank trials that followed feedback trials. The pattern of data he obtained with adult participants showed a systematic reduction of possible choices based on previous outcomes. He suggested that participants work with a limited set of hypotheses constrained by the task and that they systematically reduce this set following feedback trials. Levine argued that such systemacity requires rehearsal, which was supported by data from related experiments and postmortems with participants (Levine, 1975). This systemacity and efficiency was not observed with younger children (Gholson, Levine, & Phillips, 1972).

Later research has disconfirmed hypothesis-testing theory. Kellogg, Robbins, and Bourne (1978) used a modified version of Levine's task in which they used memory probes about the immediately preceding trial. These probes asked about features of the previous stimuli, response given, feedback received, and stated hypothesis. Recognition was high for response and feedback, but low for features and hypothesis. This is a crucial blow to Levine's interpretation of his data. Kellogg et al.'s (1978) results, though, are consistent with the rehearsal hypothesis. Recognition for hypotheses was high following positive feedback, which is when a hypothesis could be rehearsed and thus subject to extensive processing. Moreover, induced rehearsal did not improve recognition performance, consistent with Levine's suggestion that participants were spontaneously submitting themselves to rehearsal.

Overtraining through extensive processing can be implemented in cascade-correlation networks by lowering the score threshold (i.e., the allowable discrepancy between desired and actual output). With a low score-threshold, networks are trained to make finer discriminations between patterns by working more on the material. For example, the output units we typically use with cascade-correlation have a range between −0.5 and 0.5. With the default score-threshold of 0.4, output values between 0.1 and 0.5 are acceptable for target values of 0.5, and output values between −0.1 and −0.5 are accepted for target values of −0.5. With a score-threshold of 0.1, however, only output values between 0.4 and 0.5 are accepted for target values of 0.5, and only output values between −0.4 and −0.5 are acceptable for target values of −0.5. All other things equal, networks need more training trials to achieve the level of performance of a lower score-threshold. The score-threshold parameter thus affects the amount of training the networks get. In turn, amount of training increases depth of learning. The longer a network learns, the more precise its approximation to the target function defined by the training patterns.

We modeled adult behavior with a score threshold of 0.01. We used the default threshold of 0.4 to model preschool performance.[6] Spontaneous overtraining through extensive processing is thus our explanation of age differences in shift

[6] Alternatively, short of implementing a rehearsal-like module, we could specify the number of training trials, rather than a lower score-threshold, to implement overtraining. We have tested this alternative for the preshift phase, and it replicates the findings we report for postshift learning. The downside of this approach is that it fails to take into account variability between networks due to the initially random connections (i.e., some networks learn to criterion faster than others, like humans), and it prevents statistical analysis of preshift learning because variability of trials to criterion is 0. We thus preferred the score-threshold alternative. It mimics the expected effect of extensive processing by providing extra training. Adult networks will require more trials than child networks to reach criterion, but this includes extensive covert processing. It is therefore compatible with the observation that human adults require less observable trials. More epochs in adult networks, in our view, equate with extensive internal processing of external events and allow adults to solve the task quicker by external standards, even though they work harder covertly.

TABLE 1
Empirical Data

Task	Adults	Younger children
Reversal and nonreversal	Reversal easier	Reversal = nonreversal
Trial-by-trial evaluation	Impaired performance on unchanged nonreversal pairs	High performance on unchanged nonreversal pairs
Optional shift	Greater proportion of reversers	Greater proportion of nonreversers

learning, and our simulations test this hypothesis. We make no claim about the specific nature of spontaneous overtraining, e.g., whether it results from a basic and implicit process or a deliberately applied strategy. The model, though, evaluates how the spontaneous overtraining hypothesis may capture the developmental effects in discrimination shifts.

In all our simulations, we used networks of four input units. The first input unit coded shape of the left stimulus, the second color of the left stimulus, the third shape of the right stimulus, and the fourth color of the right stimulus.[7] The binary attributes of the stimuli were coded as -1 or 1. These input units were connected to two output units. Every network had initially random connection values between input and output units. This results in variability for both number of training trials to criterion and final weight configurations. Because no two networks are identical at the onset of the task, we can use samples of networks in the different experimental conditions, as we would with humans. Desired output was $[0.5, -0.5]$ when the target was at the left, and $[-0.5, 0.5]$ when it was at the right. None of the networks recruited hidden units to solve the tasks, confirming that these tasks are linearly separable.

We modeled three discrimination shift tasks using cascade-correlation networks. Simulation 1 concerns reversal and nonreversal shift tasks. Trial-by-trial data for these shifts are reported in simulation 2. Finally, simulation 3 reports data from the optional shift task. The psychological data to be captured are summarized in Table 1.

Simulation 1: Reversal and Nonreversal Shifts

In this simulation, we submitted our networks to the reversal and nonreversal shift tasks. Networks representing adults, with a low score-threshold, were expected to perform reversal shifts faster than nonreversal shifts. Networks representing preschoolers, with a higher score-threshold, were expected to learn both shifts equally easily. We also analyzed the networks' knowledge represen-

[7] We use labels such as "shape" and "color" only for clarity. Networks did not see "red circle" figures and such. The networks were only presented with vectors of numbers that were consistent with the tasks we modeled. We assume that these vectors represent preprocessed stimuli.

tations in order to better understand their behavior. Specifically, we examined plots of error over epochs, weight diagrams, principal components analyses of network contributions, and plots of output activations.

Method. One hundred twenty adult networks were used in this simulation, with the score-threshold set to 0.01. Sixty were initially trained on one attribute of color, and 60 on the other attribute of color. The irrelevant dimension was variable within trials. When performance reached threshold on all problems of the initial discrimination, training was shifted to a new attribute. In each subset of 60 networks, one-third of the networks had to learn to respond to the other attribute of the color dimension (reversal shift, $n = 20$), while the remaining networks were required to learn one of the two attributes of the shape dimension (nonreversal shift, $n = 20$ per attribute), for a total of six groups. Learning continued until criterion was reached (i.e., output activations were within threshold of the desired values for both output units on all problems). One hundred twenty child networks were used under the same conditions, with the score-threshold set at 0.4.

To assess whether networks perform faster on reversal or nonreversal shifts, we recorded the number of epochs required to reach criterion in shift learning. An epoch is, in this case, a block of four trials, one with each stimulus pair. For control purposes, we also recorded epochs to criterion for the initial learning phase. Output errors (i.e., the discrepancy between target and actual output activations) were recorded during initial and shift learning.

Weight diagrams can be useful graphical representations of connection weights in networks. These representations can inform us about how the different properties of the input contribute to the overt behavior of the networks. We recorded weight values in both types of networks (adult and child) for both types of shift (reversal and nonreversal) at the end of each training phase in order to plot such diagrams. Raw weight values were saved in a file during the computer simulations.

Another useful technique for probing into the behavior of neural networks is contribution analysis. Contributions are the products of sending unit activations and connection weights going into output units. A contribution is therefore the activity measured on one output weight on a given trial. When we record these contributions for all input patterns after training, we have a pattern by contribution matrix with which we can perform a Principal Components Analysis (PCA). A PCA, which is highly similar to a Factor Analysis, identifies the major dimensions of variation in an array of data. The factor loadings from the PCA, after standardization, are used to identify the major dimensions of the network's representation of the task.

Finally, plots of output activation can be informative when inquiring about the responses of networks to a range of input values. We recorded, at the end of initial training, output activations as a function of a range of values on both of the input units of the same side (e.g., the two input units coding attributes of the left

TABLE 2
Epochs to Criterion for Initial and Shift Phases: Adult Networks

Dimension attribute			Phase	
Initial	Shift	Type	Initial	Shift
Red	Green	Reversal	10.3	8.2
Red	Square	Nonreversal	10.4	11.7
Red	Circle	Nonreversal	10.5	12.1
Green	Red	Reversal	10.5	8.4
Green	Square	Nonreversal	10.5	12.0
Green	Circle	Nonreversal	10.3	11.8

stimulus for activations of the left output unit). The input values we used ranged between -1 and 1 (the two values used for training), with increments of 0.1. Activation of the remaining input units was set to zero.

We analyzed the adult and child networks separately because we were only concerned with the learning pattern in each group. The use of a different score-threshold and the assumption of extensive processing in adult networks prevent combining both groups in analyses of epochs to criterion. Networks learn to a higher score-threshold quicker than to a lower score-threshold. But with respect to human learning, we assume that a portion of training in adult networks corresponds to covert extensive processing. Thus comparing learning speed between adult and child networks would not be appropriate. Only for contribution analysis will differences between both types of networks be statistically evaluated, because this is a measure of the networks' representations of the problem and as such is a meaningful comparison. Comparisons of output activations and weight configurations should also be meaningful, for the same reason.

Results. The fourth column of Table 2 presents the number of epochs required by adult networks to learn the initial discrimination in each of the six groups. The results of a one-way analysis of variance show that there were no significant differences between any of the groups, $F(5,114) = .791$, n.s. The average number of epochs required to learn the initial discriminations was 10.4 ($SD = 0.53$).

Presented in the last column of Table 2 are the numbers of epochs required by adult networks to learn the shifted discrimination. The one-way analysis of variance shows that there was a significant difference between the groups, $F(5,114) = 22.99, p < .001$. The only significant differences found using Scheffé post-hoc comparisons were all those between reversal and nonreversal groups (absolute mean differences between 3.35 and 3.9). The average number of epochs to learn a reversal shift was 8.28 ($SD = 1.79$), whereas it was 11.89 for an nonreversal shift ($SD = 1.68$).

The numbers of epochs required by child networks to learn the initial discrimination are presented in the fourth column of Table 3. The results of a one-way

TABLE 3
Epochs to Criterion for Initial and Shift Phases: Child Networks

Dimension attribute			Phase	
Initial	Shift	Type	Initial	Shift
Red	Green	Reversal	4.85	4.35
Red	Square	Nonreversal	4.80	4.35
Red	Circle	Nonreversal	4.65	4.25
Green	Red	Reversal	4.65	4.20
Green	Square	Nonreversal	4.75	4.50
Green	Circle	Nonreversal	4.80	4.30

analysis of variance show that there were no significant differences between any of the groups, $F(5,114) = .463$, n.s. The average number of epochs required to learn the initial discriminations was 4.73 ($SD = 0.53$).

Presented in the last column of Table 3 are the numbers of epochs required by the child networks to learn the shifts. The one-way analysis of variance shows that there was no significant difference between the groups, $F(5,114) = .737$, n.s. The average number of epochs to learn a reversal shift was 4.23 ($SD = 0.42$), and it was 4.38 for a nonreversal shift ($SD = 0.58$). Note that child networks require less training trials than adult networks. This is expected, as mentioned in footnote 6, because the assumption is that adults perform extensive processing. We elaborate on this in the discussion.

Figure 7 shows the error curves for reversal and nonreversal shift tasks over epochs in adult networks. Error is the squared discrepancy between target and

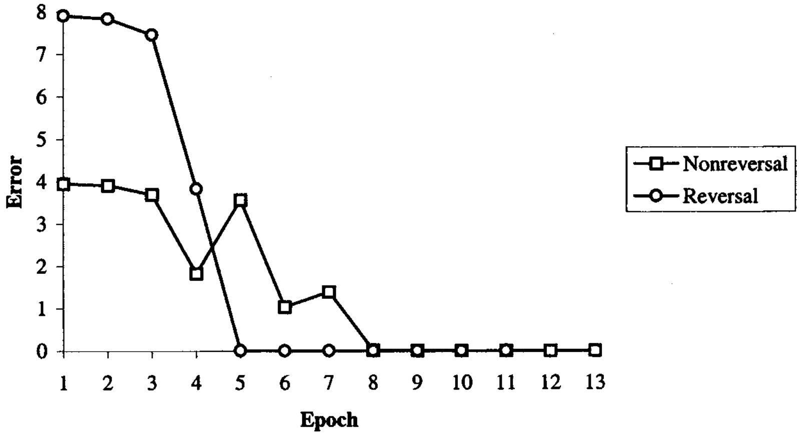


FIG. 7. Average error over epochs in shift learning: adult networks.

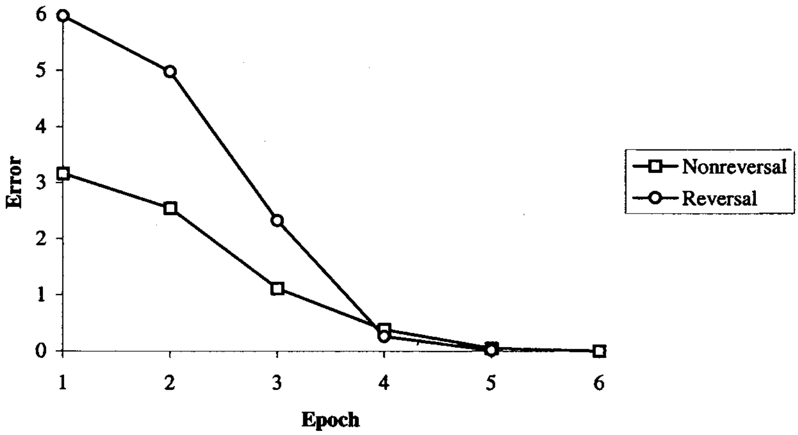


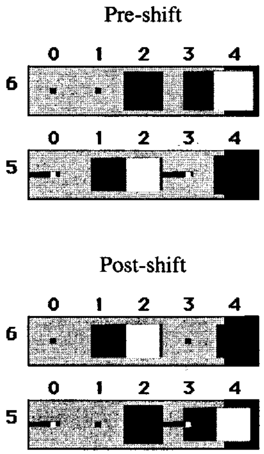
FIG. 8. Average error over epochs in shift learning: child networks.

actual output, summed over all pairs. Error for reversal shifts is initially larger and exhibits faster resolution than in the nonreversal condition. The average error at the onset of the shift training phase for adult networks is 7.90 for a reversal shift and 3.95 for a nonreversal shift. In Fig. 8, error curves for child networks are plotted. The larger initial error on reversal shifts allows for a steeper slope over epochs for this condition, but it is not sufficient to give reversal shift networks a speed advantage over their nonreversal counterparts. For child networks, average initial error is 5.97 for reversal shifts and 3.16 for nonreversal shifts.

Weight diagrams of four representative networks are presented in Fig. 9 through 12. The gray strips, labeled 5 and 6, represent the left and right output units, respectively. Within these, incoming weights are represented by squares labeled 0 (bias unit), 1 and 2 (left stimulus), and 3 and 4 (right stimulus). The size of these squares is an index of the absolute size of the standardized values of the weights. Negative (inhibitory) and positive (excitatory) weights are black and white, respectively.

For adult networks performing a reversal (Fig. 9) or a nonreversal shift (Fig. 10), we can see specific encoding of the relevant dimension. In both examples, initial training involved a target value coded on input units 2 and 4. Weights from these units have the largest values, while weights from the remaining units have the smallest values. For the network learning a reversal, shift training involves a target coded on the same units. As can be seen, the sign of each relevant weight has been reversed (bottom of Fig. 9). The nonreversal shift network, on the other hand, learned to ignore the previously relevant dimension after the shift (bottom of Fig. 10). The newly relevant dimension, coded on input units 1 and 3, is associated with the largest weights.

In child networks, for both the reversal and the nonreversal shift conditions

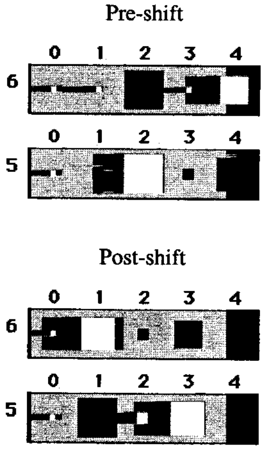

Pre-shift

Post-shift

FIG. 9. Connection weights in an adult network at the end of pre- and postshift phases: reversal shift.

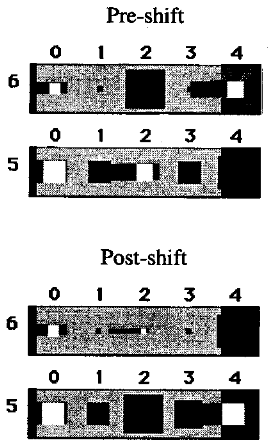
(Figs. 11 and 12, respectively), the relationship between weight configuration and relevant dimension is not as obvious as in adult networks. The weights suggest compound encoding of the stimuli rather than specific encoding of the target dimension. That is, all attributes, irrespective of their relevance to correct responses, contribute to the behavior of the networks.

Tables 4 through 7 present the standardized loadings from contribution anal-


Pre-shift

Post-shift

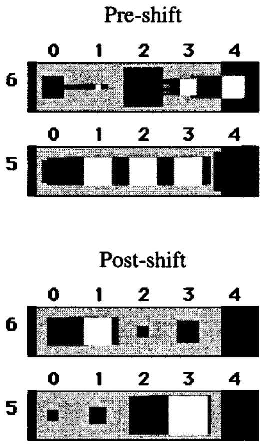
FIG. 10. Connection weights in an adult network at the end of pre- and postshift phases: nonreversal shift.


Pre-shift

Post-shift

FIG. 11. Connection weights in a child network at the end of pre- and postshift phases: reversal shift.

ysis of four representative networks. Tables 4 and 5 present loadings for two adult networks performing a reversal and a nonreversal shift, respectively, while Tables 6 and 7 present the same data for child networks. We removed loadings with zero values from these tables. The weights are indexed by their sending input unit (I1 through I4) and by their receiving output unit (O1 or O2). For the networks reported in Tables 4 through 7, initial training involved discriminating values coded on input units 2 and 4. Although this remained true for reversal shift


Pre-shift

Post-shift

FIG. 12. Connection weights in a child network at the end of pre- and postshift phases: nonreversal shift.

TABLE 4
Standardized Loadings from Contribution Analysis: Adult Reversal Shift

	Preshift		Postshift	
Weight	C1	C2	C1	C2
I1O1		1.00		1.00
I2O1	1.00		−.99	
I3O1		−.99		−.99
I4O1	.99		−1.00	
I1O2		−1.00		−.99
I2O2	−1.00		1.00	
I3O2		.99		1.00
I4O2	−1.00		1.00	

networks during shift training, the shift was to input units 1 and 3 for a nonreversal shift.

The analyses revealed a two-component solution for each network. In all child and adult networks we examined, all nonzero loadings for the first component were associated with the relevant input units; all nonzero loadings for the second component were associated with the irrelevant input units. This can be seen in Tables 4 through 7. We can refer to the first component as the "Relevant Information" component and to the second as the "Irrelevant Information" component. We observed this for both initial and shift tasks. After a RS, only the signs of the loadings are changed, whereas after a NS, loadings are switched between components such that weights previously associated with one component are now associated with the other. For a sample of 20 adult networks, the average percentage of variance explained by the first component was 96.83

TABLE 5
Standardized Loadings from Contribution Analysis: Adult Nonreversal Shift

	Preshift		Postshift	
Weight	C1	C2	C1	C2
I1O1		−1.00	−.99	
I2O1	−.99			−1.00
I3O1		1.00	−1.00	
I4O1	−1.00			−1.00
I1O2		−1.00	1.00	
I2O2	.99			−1.00
I3O2		1.00	1.00	
I4O2	.99			.99

TABLE 6
Standardized Loadings from Contribution Analysis: Child Reversal Shift

	Preshift		Postshift	
Weight	C1	C2	C1	C2
I1O1		.99		1.00
I2O1	−1.00		−1.00	
I3O1		−1.00		−.99
I4O1	−.99		−.99	
I1O2		1.00		1.00
I2O2	.99		1.00	
I3O2		−1.00		−.99
I4O2	.99		1.00	

(*SD* = 1.68). The remaining variance was accounted for by the second component (*M* = 3.17, *SD* = 1.68). For an equivalent sample of 20 child networks, the first component accounted for an average of 85.31% of the variance (*SD* = 11.43). The second component accounted for the remaining variance (*M* = 14.69, *SD* = 11.43).

Because the PCAs reveal a two-component solution that captures all variance, which was expected considering there are only two dimensions of variation in the input, any variance not accounted for by the first component is accounted for by the second component. As such, the second component is linearly determined by the first in both adult and child networks. This is seen in the standard deviations, which are identical for both components within a group of networks. The extent to which a network attends to the relevant information determines what is left for irrelevant information.

TABLE 7
Standardized Loadings from Contribution Analysis: Child Nonreversal Shift

	Preshift		Postshift	
Weight	C1	C2	C1	C2
I1O1		−.99	1.00	
I2O1	1.00			.99
I3O1		.99	.99	
I4O1	1.00			.99
I1O2		−1.00	−1.00	
I2O2	−1.00			.99
I3O2		1.00	−.99	
I4O2	−1.00			−1.00

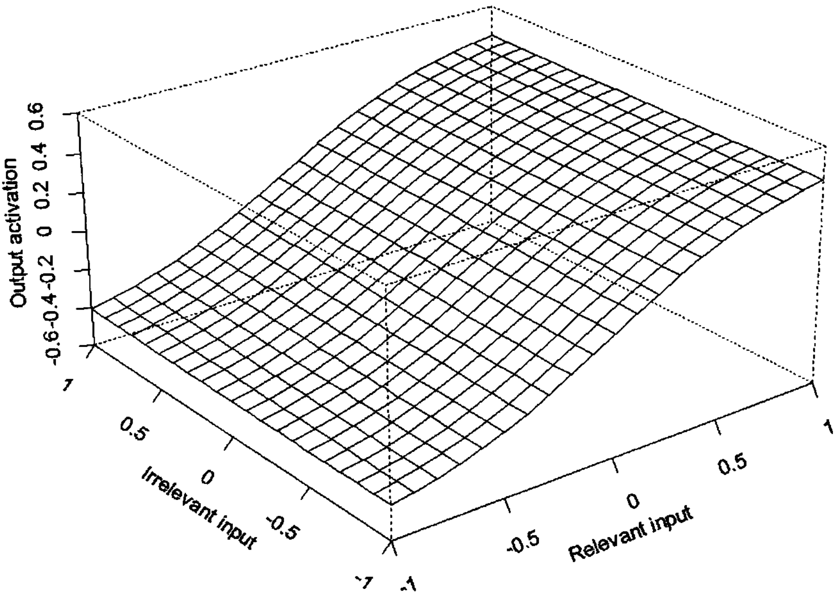


FIG. 13. Activation of the left output unit in an adult network after initial training as a function of relevant and irrelevant inputs.

We tested the difference between the amounts of variance accounted for in both types of networks in order to evaluate if the relationship between their use of relevant and irrelevant information differs. More variance is accounted for in adult networks than in child networks by the first component ($t(38) = 4.462, p < .001$), while the reverse is true for the second component ($t(38) = -4.466, p < .001$). In child networks, then, irrelevant information contributes to a larger extent to overt behavior than it does in adult networks.

Finally, Fig. 13 presents a plot of output activation as a function of input from two dimensions in a representative adult network after initial training. In this example, activation is from the output unit representing left, and both input units represent the left stimulus. The network was trained to turn this unit on when the value on the relevant dimension of the left stimulus was 1.0, and turn it off when it was -1.0, irrespective of input on the other dimension. The shape of the sigmoid activation function can be seen along the axis of the input unit coding the relevant dimension. Further, the different values along the irrelevant dimension have no observable effect on the function. In Fig. 14, the same types of data are plotted for a representative child network. The sigmoid function is difficult to identify along the relevant dimension, and the output is slightly affected by input from the irrelevant dimension. Figure 15 presents a plot for another representative child network. In this case, the output unit is responding more to the irrelevant input unit than to the relevant one.

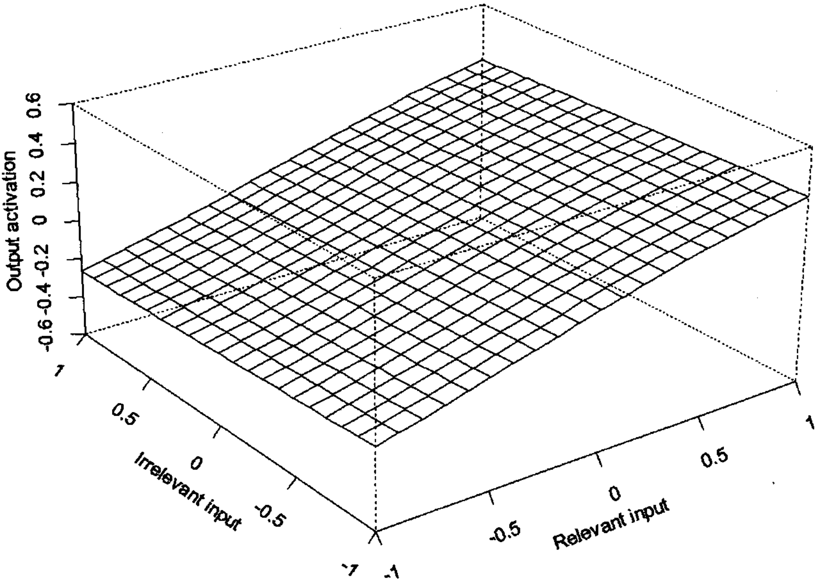


FIG. 14. Activation of the left output unit in a child network after initial training as a function of relevant and irrelevant inputs.

Discussion. Our adult networks, like human adults, performed a reversal shift faster than a nonreversal shift. This was accomplished by two-layer networks lacking any mediating units and marks a departure from both mediational accounts. Child networks, on the other hand, learned both shifts equally easily, which is consistent with the empirical data from younger children. This is thus also a departure from Tighe and Tighe's (1966a) perceptual differentiation account.

Child networks required fewer epochs to learn the initial and shift discriminations than adult networks, which was expected given the score-thresholds used. This is not the case for humans, as adults require fewer learning trials than younger children on these tasks (e.g., Wolff, 1967). But our assumption is that adults submit themselves to a form of overtraining through extensive processing. Consequently, the number of epochs to criterion cannot be equated with training trials in human experiments, which assess only the number of stimulus presentations, not the amount of processing. In our simulations, the number of epochs is an index of the amount of processing, which we use to assess the ease of the different shifts within each group of networks (child and adult).

All three network analysis techniques show that knowledge representations in adult networks at the end of the preshift phase were more precisely focused on the relevant dimension than were those of the child networks. This precise preshift discrimination enabled faster learning of a reversal shift than of a

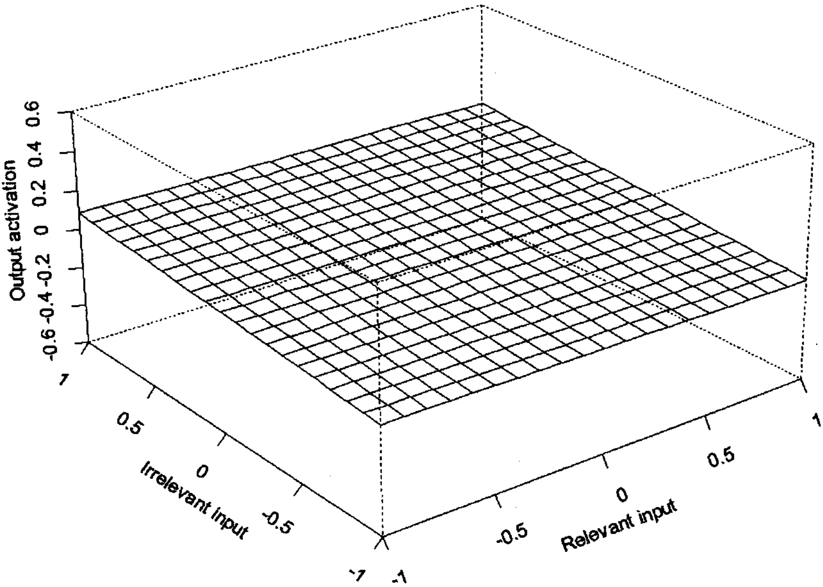


FIG. 15. Activation of the left output unit in another child network after initial training as a function of relevant and irrelevant inputs.

nonreversal shift in adult networks. In contrast, the use of irrelevant information to solve the preshift task in child networks enabled faster learning of a nonreversal shift to the point that it was as easy to learn as a reversal shift. The greater focus on the relevant dimension in adult networks can be traced to their greater amount of learning.

These distinct knowledge representations between preschool and adult networks are consistent with the verbatim-to-gist ontogeny observed in memory research (Brainerd & Reyna, 1993). Within the same age range that is associated with shift learning development (i.e., 4 to 10 years of age), children change from heavily relying on verbatim representations of inputs in problem solving to using the gist (or meaning) of the inputs as they grow older. Adult networks extract the gist of the information by ignoring irrelevant information, whereas child networks rely more on the verbatim, or compound, properties of the stimuli.

We further suggest that the compound encoding of stimulus attributes in children is not only at the level of the individual stimulus (e.g., Kendler, 1983), but that both stimuli in a pair contribute to a given response. This is exemplified in Fig. 15, where input from the right-hand stimulus would be required to produce a "left side" answer when the target attribute is associated with the left stimulus.[8] If children do indeed encode properties of the pair rather than only those of the individual stimuli, our model predicts that children will perform poorly or at chance level when asked to classify the individual stimuli as "target" or "nontarget" at the end of pairwise training. This prediction is also consistent with the verbatim-to-gist ontogeny (Brainerd & Reyna, 1993), because it assumes that younger children did not extract the relevant information from the compound (or verbatim) array of stimuli.

Before addressing other theoretical implications of these findings, we turn to the trial-by-trial performance of both types of networks on the different stimulus pairs during shift training.

Simulation 2: Trial-by-Trial Analyses

In a nonreversal shift, correct responses do not change for half of the problems after initial training. Preschoolers exhibit high performance on these unchanged pairs throughout shift learning. Yet for most older children and adults, there is a drop in performance for those unchanged pairs. This simulation tested the ability of child and adult networks to exhibit these distinct behaviors.

Method. Sixty adult networks were used in this simulation, with the training score-threshold set at 0.01. All were trained with the same initial attribute of color as target. The irrelevant dimension was variable within trials. When criterion was reached, 20 networks were trained on a reversal shift, and the remaining networks were trained on a nonreversal shift ($n = 20$ per each of two attributes of shape). Throughout the shift learning phase, we recorded the proportion of pairs for which networks responded correctly. A network was correct when activations on both output units were within 0.4 of the target values. We used a testing criterion of 0.4 because this is the default score-threshold, which we use to train child networks. We therefore could not use a lower value to appropriately evaluate child networks, and we wanted the assessment of performance for both types of network to be identical. Sixty child networks were also used in this simulation, under the same conditions, with the training score-threshold set at 0.4.

Results. Figure 16 shows the proportion of pairs to which adult networks respond correctly over epochs. Performance on reversal and changed nonreversal pairs increases over epochs. Performance on unchanged nonreversal pairs is initially high, but drops to zero when performance increases on the changed pairs. When performance on unchanged pairs improves, there is a drop in performance on changed pairs. Performance on changed pairs increases again and results on both types of pairs remain high until criterion.

For child networks, performance on the unchanged pairs remains high throughout shift learning (Fig. 17), while performances on reversal and changed

[8] It should be mentioned that both stimuli also contribute to both output activations in adult networks. But if one is removed, the output is little affected, as can be seen in Fig. 11. That is, an adult network correctly identifies a stimulus as target or nontarget even when the other pair member is not presented.

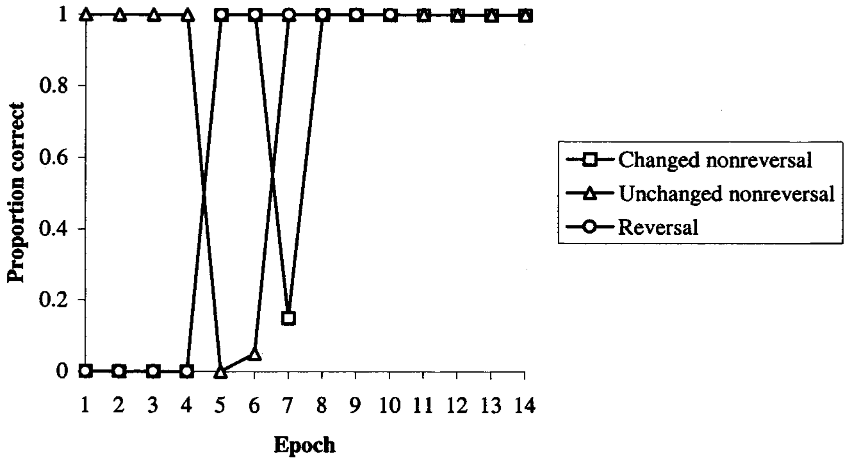


FIG. 16. Proportion of correct responses during shift training: adult networks.

nonreversal patterns increase at similar rates. In Fig. 18, we present child data from Tighe and Tighe (1978) on the same tasks.

Discussion. Our adult networks exhibit behavior that mirrors human adult data. Although the reward contingencies do not change for half of the stimulus pairs in a nonreversal shift, human adult performance on these pairs drops and then increases before criterion is reached (Kendler, 1979; Tighe & Tighe, 1978). We observe this same pattern for unchanged nonreversal pairs in our adult networks.

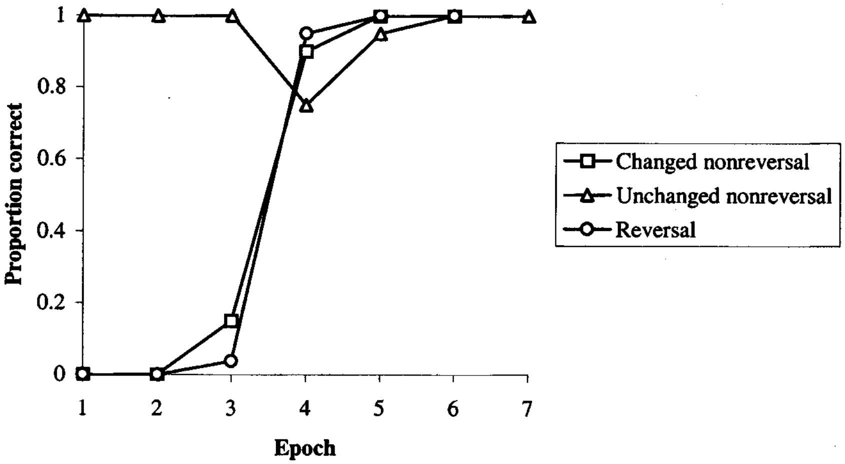


FIG. 17. Proportion of correct responses during shift training: child networks.

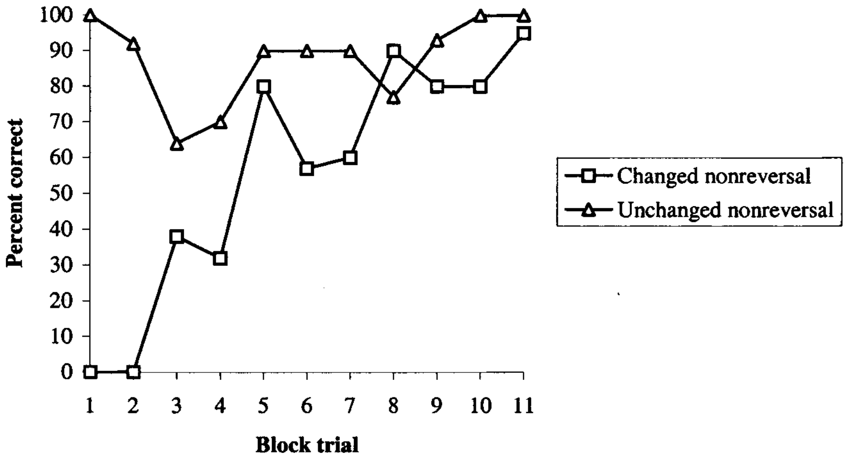


FIG. 18. Proportion of correct responses over trials during nonreversal shift training in human preschoolers (reprinted, with permission, from Tighe & Tighe, 1978).

It should be noted again that this pattern of behavior is not the result of intermediate processes. Previous nonreversal shift interpretations suggested that mediating participants initially shift their overt responses to the original dimension response, which impairs performance on the unchanged pairs (Kendler, 1979; Kendler & Kendler, 1969). Learning a nonreversal shift thus involved the extinction of the mediated response to the initial dimension and retraining on the newly relevant one.

In our adult networks, a nonreversal has the same initial deleterious effect on the performance of the unchanged patterns. This is because the finely tuned discriminations learned during the preshift phase must be focused onto another dimension. However, our networks generate this effect without any mediating hidden units. This result is consistent with Tighe and Tighe's (1966a) perceptual differentiation model.

The importance of fine tuning is emphasized by the behavior of our preschool networks, for which performance on unchanged pairs remains high throughout shift training, as it does with human preschoolers. Although these networks are presented with the same task as the adult ones, we do not observe the same initial deleterious effect of the nonreversal shift on the unchanged patterns. For these networks, not ignoring the irrelevant dimension during initial training allows them to maintain adequate performance on the unchanged patterns during shift training. At the onset of the nonreversal shift, the child networks are responding to a greater extent than adult networks to input from the newly relevant dimension, preventing deterioration of performance for unchanged patterns. This is also consistent with Tighe and Tighe's (1966a) interpretation. They suggested that adults differentiate the perceptual properties of the stimuli more finely than children and explained the developmental differences on nonreversal shift pairs as we do here (Tighe & Tighe, 1978).

We next report our simulation of the optional shift task, which has been used extensively to differentiate associative and concept-mediated behavior (Kendler, 1983, 1995).

Simulation 3: Optional Shifts

In almost all older children and adults, performance on the test pairs of the optional shift is consistent with a reversal shift. Our adult networks are thus expected to be reversers. Younger children, though, do not exhibit reversal behavior. We expect our child networks to be nonreversers.

Method. Sixty adult networks were used in this simulation, with the score-threshold set at 0.01. Initial training was on one attribute of color for 30 networks and on the other attribute of color for the remaining 30 networks. The irrelevant dimension was variable within trials. When criterion was reached, the networks were trained to shift their responses on two of the four stimulus pairs. This shift was congruent with both a reversal and a nonreversal shift. When training was completed in this second phase, we recorded the networks' behavior on the remaining two pairs. Sixty child networks were trained and tested in the same conditions, with the score-threshold set at 0.4.

As in the psychological literature, an individual network was labeled reverser only if its performance on both test pairs was consistent with a reversal shift; otherwise, it was labeled nonreverser. Response to a specific pair can be "left," "right," or "guess" (if both output units are on or both are off, we assume that choosing only one of the two stimuli, a requirement of the task, is equivalent to a random process). In the case of a guess response, we assign a .5 probability of a reversal for that pair. That is, a random choice of either of the two stimuli, where only one is consistent with a reversal shift, results in a 50% chance of choosing it.

A network that chose the reversal stimuli in both pairs was assigned a probability of 1 to be labeled reverser. When a network chose the reversal stimulus in one pair but guessed for the other, it was assigned a probability of .5 to be labeled a reverser. And if a network guessed on both test pairs, there was a probability of .25 of being labeled reverser. Choosing the stimulus associated with a nonreversal in any of the two test pairs resulted in a probability of 0 to be labeled reverser. For adult and child networks, we tabulated these probabilities in order to evaluate how many networks could be expected to perform as reversers. The remaining networks are labeled nonreversers. These numbers are compared to the empirical distribution of reversers and nonreversers in humans.

Results. Fifty-seven adult networks were labeled as reversers, whereas the behaviors of the remaining three were consistent with a nonreversal shift on both pairs. No network produced a guess answer for any pair. Table 8 presents the percentage of reverser and nonreverser adult networks, as well as human data

TABLE 8
Percentage of Reversers and Nonreversers in Adult Networks and Human Adults

	Reversers (%)	Nonreversers (%)
Networks	95	5
Humans[a]	87	13

[a] Human data from Kendler (1983).

reported in Kendler (1983). A χ^2 goodness-of-fit test indicates no significant difference between our network results and those for adults ($\chi^2(1, N = 60) = 3.39$, n.s.).

Thirty-six of sixty child networks showed behavior consistent with a nonreversal shift on both test pairs (nonreversers). Five guessed on one pattern but produced nonreversal behavior on the other (nonreversers as well). One network produced reversal behavior on both pairs, four guessed on one pair but produced reversal behavior on the other (.5 × 4: probability of 2 reversers), and 14 networks had to guess on both pairs (.25 × 14: probability of 3.5 reversers), for an overall probability of 6.5 reversers in our 60 child networks. Table 9 presents the percentages of reverser and nonreverser child networks, as well as human data reported in Kendler (1983). A χ^2 goodness-of-fit test indicates no significant difference between our network results and those for children ($\chi^2(1, N = 60) = 3.15$, n.s.).

Discussion. Most of our adult networks, like human adults, shifted their initial responses on the test pairs. That is, their behavior was consistent with a reversal shift. In the case of our child networks, most exhibited behavior not consistent with a reversal shift.

The finely tuned discriminations learned by adult networks, but not child networks, in the initial discrimination explain these differences. For adult networks, focusing on the initially relevant dimension makes the error signal associated with input units from that dimension stronger than the error associated with the other input units. Because learning in cascade-correlation networks involves reducing the largest source of error, weight changes after the shift

TABLE 9
Percentage of Reversers and Nonreversers in Child Networks and Human Preschoolers

	Reversers (%)	Nonreversers (%)
Networks	10.83	89.17
Humans[a]	20	80

[a] Human Data from Kendler (1983).

concentrate on the initial dimension, thus making the initial responses to the test pairs consistent with a reversal shift.

In child networks, initial training does not involve ignoring the irrelevant dimension. When the shift phase begins, changes in weights are made for both dimensions to accommodate the new contingencies for the shift pairs. For a large proportion of child networks, these changes do not significantly alter the compound responses previously learned for what are now the test pairs, as was the case for the unchanged pairs of the nonreversal shift for which performance remained high. Indeed, the test pairs in an optional shift are the same as the unchanged pairs of an nonreversal shift would be (e.g., in Fig. 3, a nonreversal shift from *square* to *green* after initial training would involve no change for the pairs that have the "green square" stimulus, which are the test pairs in that optional shift example).

GENERAL DISCUSSION

Our simulations of reversal, nonreversal, and optional shifts were designed according to the assumption that older children and adults, compared to kindergarten children, learn finer tuned discriminations from a process similar to rehearsal in memory tasks, a suggestion made over 30 years ago by Levine (1966, 1975). We did not build in mediating mechanisms to model adult behavior (Raijmakers et al., 1996). Our model captures the empirical data and suggests a new explanation of human shift learning.

In the introduction, we emphasized that developmental differences in discrimination shift learning between younger and older children are observed. Contrary to the pervasive belief that nonreversal shifts are easier than reversal shifts for younger children, though, we suggested that equal ease of reversal and nonreversal shifts is a more accurate reflection of the evidence. We also presented the limitations of the previous theoretical accounts of shift learning and the consequent lack of a general and thorough account of the empirical data. What, then, is the contribution of the work we have presented in this paper?

Unlike the classic explanations that mature performance on these tasks requires mediated processing (Kendler, 1983; Kendler & Kendler, 1962, 1969; Zeaman & House, 1974), our networks do not have intermediate units to represent concepts or attentional responses to modulate overt behavior. Conceptual or attentional mediation would be implemented by hidden units, which are not required by neural networks for these linearly separable problems. Our neural network model is closer to the perceptual model of Tighe and Tighe (1966a, 1978). Adult networks do indeed differentiate between relevant and irrelevant information better than child networks. But our model accounts for the equal ease of nonreversal and reversal shifts in preschoolers, whereas Tighe and Tighe's predicts easier nonreversal shifts. Table 10 summarizes the ability of all three theories and of the spontaneous overtraining model to account for shift learning regularities to which they all apply in continuous tasks. The spontaneous over-

TABLE 10
Psychological Regularities Captured by the Different Models in Continuous Paradigms

	Model			
Regularity	Conceptual mediation	Attentional mediation	Perceptual differentiation	Spontaneous overtraining
Adult reversal < nonreversal	Yes	Yes	Yes	Yes
Adult impaired performance on unchanged pairs	Yes	Yes	Yes	Yes
Adult optional shift reversal	Yes	Yes	Yes	Yes
Child reversal = nonreversal	No	No	No	Yes
Child unimpaired performance on unchanged pairs	Yes	No	Yes	Yes
Child optional shift nonreversal	Yes	No	Yes	Yes

training model clearly provides the best coverage of the psychological regularities at this point in time.

Raijmakers et al. (1996) concluded that neural network learning should be described as merely associative. Although their networks used hidden units, our networks instead use direct connections between input and output. Yet our networks still capture learning phenomena that are believed to be rule-governed, rather than merely associative. The initial goal of this project was to take on the challenge posed by Raijmakers and her colleagues (1996). We feel that we have presented a model that meets that challenge. Neural network models can capture the apparent rule-governed nature of human learning in a principled way that suggests novel theoretical interpretations and empirical predictions. From these simulations, we predict poor classification performance in younger children after pairwise discrimination training. This prediction stems from the observation that the behavior of child networks is essentially a function of the compound properties of the pair of stimuli, and not of the individual stimuli.

Kruschke (1996) modeled a categorization task with a connectionist architecture called AMBRY. He obtained promising results in simulating compound categorization shifts that qualitatively capture human data. The target categories in compound categorization are logical relationships between the different dimensions (e.g., “large” or “white” but not both, a form of exclusive-or), and the stimuli are presented individually. Two key elements of the AMBRY model are the attentional units, which learn to target the relevant dimensions and in turn affect learning of the appropriate responses, and the mediating internal categories. The model learns what to attend to and how to act upon elicited categories. This is similar to the theoretical model advocated by Zeaman and House (1963, 1974), where learning is also mediated by attentional responses. Our model differs in that attention is implicit to learning. Cascade-correlation networks learn by giving greater attention to what is relevant by increasing connection weights from relevant inputs. Which model can be better at providing a general account of concept-shift tasks will be resolved only through further research.[9]

In conclusion, the model presented in this paper offers a promising alternative to previous theoretical interpretations. Rather than attributing developmental differences in discrimination shift learning to conceptual (e.g., Kendler, 1983) or attentional mediation (e.g., Zeaman & House, 1963) or to perceptual experience (e.g., Tighe & Tighe, 1966a), our account emphasizes the extensive learning afforded by spontaneous overtraining. This is consistent with the overtraining literature, which is not the case for the mediational theories. Developmental improvements in the amount of spontaneous processing, through the ages of 4 to 10, affect performance on discrimination shifts by ensuring focused learning.

We make this suggestion by applying a domain-general model to a specific task. For example, the absence of hidden units in our model is not a design decision based on assumptions with respect to the task. It is a consequence of applying the general-purpose cascade-correlation algorithm to discrimination shifts. Generative algorithms are believed to be appropriate to simulate cognitive development (Mareschal & Shultz, 1996; Quartz, 1993; Shultz & Mareschal, 1997). In the case of concept-shift tasks, we show the potential of generative algorithms for modeling learning phenomena as well. Our model is not a theory of human learning, but rather a tool that highlights how shift learning tasks might be solved. The heuristic value of the model is to suggest a theoretical alternative that can be further investigated psychologically.

The modeling reported in this paper does not rule out mediation through intermediate representations in adults and older children. It could turn out that the model is inadequate, incomplete, or incorrect. At the moment though, the model seems worth considering because it offers the best current coverage of the phenomena in the literature. The model suggests that cognitive change in discrimination learning could be a function of age-related change in the amount

[9] The task used by Kruschke (1996) is, from a modeling perspective, linearly inseparable. We thus expect cascade-correlation networks, when trained on it, to recruit hidden units that may act like mediating categories. Because the weights coming into these hidden units are frozen after their recruitment, we expect the networks to exhibit the dimensional perseveration that Kruschke observed at the onset of a shift (making a reversal easier than other shifts). On the other hand, mediation is an explicit architectural property of the AMBRY model (as it was in Raijmakers' networks). How AMBRY would perform on the linearly separable tasks we have presented might be a crucial test of mediational theory, as we have shown that mediation is not required in order to capture adult behavior in discrimination shifts.

of processing on discrete trials. Because the model ignores how spontaneous overtraining may be achieved in humans, its worth remains an empirical question. However, it is consistent with the literature and suggests a novel alternative to account for the shift learning ontogeny.

The simulations reported here represent an initial attempt to provide a theoretical account of human shift learning with the use of neural network tools. More work is still required. For example, we need to determine if our results can be replicated with other learning algorithms. Also, simulations of other concept-shift tasks would be useful to evaluate the generality of our model. Of special interest would be to capture Zeaman and House's (1974) finding that young children execute an intradimensional shift faster than an extradimensional shift when new stimuli are introduced at the onset of the shift. We trained adult and preschool networks on these tasks. Our results show intradimensional superiority over extradimensional for both child and adult networks (Sirois & Shultz, 1998).

Variability of the ease of reversal and nonreversal shifts within age levels and within individual networks should also be further examined. Cole's (1973, 1976) observation that young children will, on probe trials or at the onset of a shift, alternatively exhibit behavior consistent with a reversal or a nonreversal shift (or consistent with concept or instance responding, respectively) is not an unreasonable expectation for our child networks. Given that their connection weights are responding to both dimensions, it is possible that for some training pairs reversal shifts will be easier than nonreversal shifts, whereas the reverse would be observed for other training pairs. Performance on probe stimuli could be consistent with concept learning in one training phase (e.g., initial learning) and consistent with instance (or compound) learning in another phase (e.g., shift learning). Some of our child networks did perform the reversal shift faster than others performed the nonreversal shift. Our present results are at least consistent with Cole's. Further work should focus on this issue.

There is also the issue of dimensionless experiments (e.g., Bogartz, 1965; Goulet & Williams, 1970). Because any stimulus in such an experiment has a distinguishing feature with respect to the other stimuli, we expect adult networks to find a full reversal easier than a half reversal simply because they will focus on these features, given a strict score-threshold, as they did in the simulations reported here. Alternatively, child networks are expected to find both shifts as equally easy because in our simulations child networks did not appear to respond exclusively to the relevant dimension. This suggestion will obviously require subsequent inquiry. Finally, we need to further investigate our suggestion that child networks and children themselves would perform poorly in a classification task of the individual stimuli following the pairwise learning task.

The greater purpose of these simulations is to help formulate, by means of neural network tools, a psychological interpretation of discrimination learning. What we have outlined in this paper is, we hope, a contribution to this legitimate and much needed formulation.

REFERENCES

Bogartz, W. (1965). Effects of reversal and nonreversal shifts with CVC stimuli. *Journal of Verbal Learning and Behavior,* **4,** 484–488.

Brainerd, C. J., Olney, C. A., & Reyna, V. F. (1993). Optimization versus effortful processing in children's cognitive triage: Criticism, reanalyses, and new data. *Journal of Experimental Child Psychology,* **55,** 353–373.

Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review,* **100,** 42–67.

Buckingham, D., & Shultz, T. R. (1994). A connectionist model of the development of velocity, time and distance concepts. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 72–77). Hillsdale, NJ: Erlbaum.

Case, R. (1978). Intellectual development from birth to adulthood: A neo-piagetian interpretation. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 37–71). Hillsdale, NJ: Erlbaum.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology,* **33,** 386–404.

Cole, M. (1973). A developmental study of factors influencing discrimination transfer. *Journal of Experimental Child Psychology,* **16,** 126–147.

Cole, M. (1976). A probe trial procedure for the study of children's discrimination learning and transfer. *Journal of Experimental Child Psychology,* **22,** 499–510.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* **11,** 671–684.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Esposito, N. J. (1975). Review of discrimination shift learning in young children. *Psychological Bulletin,* **82,** 432–455.

Fahlman, S. E., & Lebierre, C. (1990). *The cascade-correlation learning architecture* (Tech. Rep. No. CMU-CS-90-100). Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development,* **37,** 283–299.

Gholson, B., & Schuepfer, T. (1979). Commentary on Kendler's paper: An alternative perspective. *Advances in Child Development and Behavior,* **13,** 137–144.

Gholson, B., Levine, M., & Phillips, S. (1972). Hypotheses, strategies, and stereotypes in discrimination learning. *Journal of Experimental Child Psychology,* **13,** 423–446.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review,* **62,** 32–41.

Goulet, L. R., & Williams, K. G. (1970). Children's shift performance in the absence of dimensionality and a learned representational response. *Journal of Experimental Child Psychology,* **10,** 287–294.

Hagen, J. W., Hargrove, S., & Ross, W. (1973). Prompting and rehearsal in short-term memory. *Child Development,* **44,** 201–204.

Hagen, J. W., Jongeward, R. H., & Kail, R. V. (1979). Cognitive perspectives on the development of memory. In A. Floyd (Ed.), *Cognitive development in the school years* (pp. 129–161). New York, NY: Halsted.

Inglis, J., Ankus, M. N., & Sykes, D. H. (1968). Age-related differences in learning and short-term-memory from childhood to the senium. *Human Development,* **11,** 42–52.

Kellogg, R. T., Robbins, D. W., & Bourne, L. E., Jr. (1978). Memory for intratrial events in feature identification. *Journal of Experimental Psychology: Human Learning and Memory,* **4,** 256–265.

Kendler, T. S., & Kendler, H. H. (1959). Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology,* **58,** 56–60.

Kendler, H. H., & Kendler, T. S. (1962). Vertical and horizontal processes in problem solving. *Psychological Review,* **69,** 1–16.

Kendler, H. H., & Kendler, T. S. (1969). Reversal-shift behavior: Some basic issues. *Psychological Bulletin,* **72,** 229–232.

Kendler, H. H., & Kendler, T. S. (1975). From discrimination learning to cognitive development: A neobehavioristic odyssey. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 1, pp. 191–247). Hillsdale, NJ: Erlbaum.

Kendler, T. S. (1979). The development of discrimination learning: A levels-of-functioning explanation. *Advances in Child Development and Behavior,* **13,** 83–117.

Kendler, T. S. (1983). Labeling, overtraining and levels of function. In T. J. Tighe, & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analysis* (pp. 129–162). Hillsdale, NJ: Erlbaum.

Kendler, T. S. (1995). *Levels of cognitive development.* Mahwah, NJ: Erlbaum.

Kendler, T. S., Kendler, H. H., & Wells, D. (1960). Reversal and nonreversal shifts in nursery school children. *Journal of Comparative and Physiological Psychology,* **53,** 83–88.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science,* **8,** 201–223.

Levine, M. (1966). Hypothesis behavior in humans during discrimination learning. *Journal of Experimental Psychology,* **71,** 331–338.

Levine, M. (1975). *A cognitive theory of learning: Research on hypothesis testing.* Hillsdale, NJ: Erlbaum.

Lips, H. M. (1993). *Sex and gender: An introduction* (2nd ed.). Mountain View, CA: Mayfield.

Mareschal, D., & Shultz, T. R. (1993). A connectionist model of the development of seriation. In M. Ringle (Ed.), *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 676–681). Hillsdale, NJ: Erlbaum.

Mareschal, D., & Shultz, T. R. (1996). Generative connectionist networks and constructivist cognitive development. *Cognitive Development,* **11,** 571–603.

Naus, M. J., Ornstein, P. A., & Aivano, S. (1977). Developmental changes in memory: The effects of processing time and rehearsal instructions. *Journal of Experimental Child Psychology,* **23,** 237–251.

Naus, M. J., Ornstein, P. A., & Kreshtool, K. (1977). Developmental differences in recall and recognition: The relationship between rehearsal and memory as test expectation changes. *Journal of Experimental Child Psychology,* **23,** 252–265.

Ornstein, P. A., Naus, M. J., & Liberty, C. (1975). Rehearsal and organizational processes in children's memory. *Child Development,* **46,** 818–830.

Ornstein, P. A., Naus, M. J., & Miller, T. D. (1977). The effects of list organization and rehearsal activity on children's free recall. *Child Development,* **48,** 292–295.

Quartz, S. R. (1993). Neural networks, nativism, and the plausibility of constructivism. *Cognition,* **48,** 223–242.

Raijmakers, M. E. J., van Koten, S., & Molenaar, P. C. M. (1996). On the validity of simulating stagewise development by means of PDP networks: Application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science,* **20,** 101–136.

Raijmakers, M. E. J. (1996). *Epigenesis in neural network models of cognitive development.* Unpublished doctoral dissertation, University of Amsterdam, Amsterdam.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review,* **65,** 368–408.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318–364). Cambridge, MA: MIT Press/Bradford Books.

Shultz, T. R. (1998). A computational analysis of conservation. *Developmental Science,* **1,** 103–126.

Shultz, T. R., Buckingham, D., & Oshima-Takane, Y. (1994). A connectionist model of the learning of personal pronouns in English. In S. J. Hanson, T. Petsche, M. Kearns, & R. L. Rivest (Eds.), *Computational learning theory and natural learning systems: Vol. 2. Intersection between theory and experiment* (pp. 347–362). Cambridge, MA: MIT Press.

Shultz, T. R., & Mareschal, D. (1997). Rethinking innateness, learning, and constructivism: Connectionist perspectives on development. *Cognitive Development,* **12,** 467–490.

Shultz, T. R., Mareschal, D., & Schmidt, W. C. (1994). Modeling cognitive development on balance scale phenomena. *Machine Learning,* **16,** 57–86.

Siegler, R. S. (1978). The origins of scientific reasoning. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 109–149). Hillsdale, NJ: Erlbaum.

Siegler, R. S. (1998). *Children's thinking* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Sirois, S., & Shultz, T. R. (1998). Neural network models of discrimination shifts. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 980–985). Mahwah, NJ: Erlbaum.

Tighe, L. S., & Tighe, T. J. (1966a). Discrimination learning: Two views in historical perspective. *Psychological Bulletin,* **5,** 353–370.

Tighe, T. J., & Tighe, L. S. (1966b). Overtraining and optional shift behavior in rats and children. *Journal of Comparative and Physiological Psychology,* **62,** 49–54.

Tighe, T. J., & Tighe, L. S. (1972). Reversals prior to solution of concept identification in children. *Journal of Experimental Child Psychology,* **13,** 488-501.

Tighe, T. J., & Tighe, L. S. (1978). A perceptual view of conceptual development. In R. D. Walk & H. L. Pick, Jr. (Eds.), *Perception and experience.* New York: Plenum.

Wolff, J. L. (1967). Concept-shift and discrimination-reversal learning in humans. *Psychological Bulletin,* **68(6),** 369–408.

Zeaman, D., & House, B. J. (1963). The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency* (pp. 159–223). New York: McGraw–Hill.

Zeaman, D., & House, B. J. (1974). Interpretations of developmental trends in discriminative transfer effects. In A. D. Pick (Ed.), *Minnesota Symposium in Child Psychology,* Vol. 8. Minneapolis, MN: University of Minnesota Press.

Zeaman, D., & House, B. J. (1984). Intelligence and the process of generalization. In P. H. Brooks, R. Sperber, & C. McCauley (Eds.), *Learning and cognition in the mentally retarded* (pp. 295–309). Hillsdale, NJ: Erlbaum.

Received: June 20, 1997; revised: March 10, 1998; July 24, 1998; September 14, 1998