

Failures to Learn and their Remediation: A Hebbian Account

James L. McClelland

McClelland, J. L. (2001). Failures to learn and their remediation: A Hebbian account. In J. L. McClelland and R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives*, (pp. 97-121). Mahwah, NJ: Lawrence Erlbaum Associates.

J.L. McClelland, Departments of Psychology and Computer Science, Carnegie Mellon University, and Neuroscience, University of Pittsburgh, and the Center for the Neural Basis of Cognition, 115 Mellon Institute, Pittsburgh, PA 15213.

This Carnegie Symposium celebrates a growing convergence of behavioral and neural approaches to the mechanisms of cognitive change and reflects an overall convergence of behavioral and neural approaches to all aspects of cognition and cognitive development. This article is a part of a corresponding convergence in my own research. In my own formative years as a psychologist, facts about underlying neural mechanisms were considered to be nearly irrelevant to understanding cognition and its development. Where I went to graduate school we studied cognition and cognitive development on the one hand and physiological psychology on the other, and they seemed almost completely disconnected subjects. My own early experimental work stayed strictly on the cognitive side of this huge divide.

Things began to change for me in the late 1970's, with the emergence of connectionist models. They seemed to many of us who were involved in their development to represent a clear step toward bridging the gap between mind and brain (Hinton & Anderson, 1981). We saw information processing as arising from the interactions of vast numbers of neurons. We began to think of propagation of activation among neurons via their synaptic connections as producing cognitive outcomes ranging from perception to comprehension to problem solving, and we thought of changes in the synaptic connections among the neurons as the basis for changes in these processes. Memory, learning, and development were all taken to be based on changes in the strength and distribution of synaptic connections.

This way of thinking could be traced back to 19th century physiologists and psychologists and was quite fully articulated fifty years ago by Donald Hebb (1949). It died out for a time in the 1960's but emerged again around 1980, stimulated in part by the advent of computers that were fast enough and cheap enough to allow the exploration of computer simulations of neural processes. In my own case, I found this approach appealing since it gave answers to many questions about cognition that had previously puzzled me, questions that I had not previously found answers to when considering them within other information processing frameworks (McClelland, 1979; McClelland & Rumelhart, 1981). The connectionist (Feldman & Ballard, 1982) or parallel-distributed processing (Rumelhart, McClelland, & the PDP Research Group, 1986b) framework for modeling cognitive processes arose around this approach, and several interesting and successful models of cognitive and developmental processes have been constructed within it.

However, the promise of connectionism for aiding the rapprochement between psychology and neuroscience was not immediately kept. One reason for this, in my view, was that many clever connectionists focussed more on computational than on neurobiological considerations. They wanted procedures for adjusting connection weights that would solve the hard learning problems, and they did not constrain themselves to consider only mechanisms that seemed consistent with the details of the underlying neurobiological processes. The canonical example is the back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986a). Back propagation provided a powerful solution to a problem previously viewed as intractable (Minsky & Papert, 1969) and has been at the heart of many successful cognitive and developmental models, but it did not seem to neurobiologists to be at all plausible in terms of the underlying neurobiological mechanisms (see Mazzoni, Andersen, & Jordan, 1991 for discussion). In neurobiological circles, the use of back propagation has sometimes been tolerated as a vehicle for constructing networks that solve biologically interesting problems, such as mapping from eye position to position in a head-centered coordinate system, taking position of the eyes in the head into account (Zipser & Andersen, 1988). But backprop is generally not thought to reflect the way in which adjustments to connection weights are actually carried out in the brain, and so is generally considered irrelevant to the study of the neural basis of

learning from a biological point of view. Some effort has been made to find computationally powerful algorithms that are more biologically plausible (Mazzoni et al., 1991; Hinton & McClelland, 1988; O'Reilly, 1996), but to my knowledge the psychological or functional implications of these proposals have not been very fully explored.

The key point is that in spite of all that was gained by adopting the computationally powerful backpropagation algorithm, something important was lost. It was not just the chance to sit down and have a good conversation with a neurobiologist. Rather, it was the opportunity to exploit what we do know about the neurobiology of learning in our efforts to understand learning at the behavioral level: particularly, when improvements in functions such as perception or skilled performance will occur due to experience, and when they will not occur.

The work I will describe today began with the following thought: Hebb's original proposal for how learning may occur in the brain is strongly supported by current research in neurophysiology. Perhaps this proposal can help us understand some cases in which experience can fail to lead to cognitive change, and perhaps it can also suggest procedures we can use to induce change in these cases. I hope this chapter will indicate that this thought has been a useful guide for some new research, leading to some interesting new findings, some of which support the Hebbian proposal. However, I will suggest at the end of the paper that Hebb's proposal, even if it is partially correct, may need further elaboration before it can stand as an adequate guide to the conditions under which experience will lead to improvements in functional outcomes.

I will start with two puzzling failures of cognitive change that began to intrigue me a couple of years ago. After reviewing them, I will suggest how they might be explained as arising from a paradoxical property of Hebb's proposed mechanism of learning. I will consider supporting data from the literature and new experimental and modeling studies that we have carried out to explore the issues further, particularly in the domain of inducing change in adults' perception of phonological distinctions not present in their native language.

Before I begin, I would like to acknowledge that the work has been deeply influenced by two other participants in this symposium. First, over many years Michael Merzenich's work has inspired me to think about the mechanisms of cognitive change in the brain. Merzenich has had very many things to say about the neurobiology of learning (see his chapter in this volume for a presentation of some of his ideas), and many of the things I have to say appear to me to be explicit or at least implicit in his work. Second, one of the two puzzles I am about to introduce was put to me by Helen Neville at a previous symposium. At the time I was already thinking about the first of the two puzzles, and it was the juxtaposition of the two that led me to think about the implications of Hebbian learning. So Merzenich and Neville deserve the credit (or, alternatively, the blame) for setting me off on the trajectory of this work. Many other colleagues, to whom I will refer as I go along, have also played key roles in shaping the ideas.

Two Puzzles

The first puzzle lies in the pattern of spared and impaired learning seen in individuals with amnesia arising from damage to the medial temporal lobes. Such patients have profound deficits in the ability to acquire new explicit memories for the contents of particular episodes and experiences, but show apparently normal improvements in performance in many different kinds of tasks where explicit memory of prior performance appears not to be required (See Squire, 1992 for a review). According to a theory that McNaughton, O'Reilly and I developed, amnesia results from the loss

of a system in the medial temporal lobes that is specialized for the rapid acquisition of the arbitrary conjunctions of elements that together make up an episode or specific experience. In this theory (McClelland, McNaughton, & O'Reilly, 1995), spared learning in amnesia reflects changes in the strengths of connections among neurons outside the medial temporal areas that occur in the course of carrying out information processing operations. These changes are thought to be relatively small in magnitude, so that they cannot lead to the formation of an arbitrary association in one or even a few repetitions, but they are thought to be sufficient to produce item-specific priming effects that can arise from a single processing event, and their accumulation over many processing events is thought to lead to the acquisition of cognitive skill. Furthermore, gradual accumulation of such changes through repeated activation of arbitrary association is thought to result in the establishment of these associations outside the hippocampal system, accounting for the amnesic's ability to retain arbitrary knowledge acquired well before the onset of amnesia.

If the theory of McClelland et al. (1995) is correct, it becomes puzzling that it has sometimes proven impossible to teach amnesics new arbitrary associations even with extensive repetition. A clear case in point is provided by an experiment by Gabrieli, Cohen, and Corkin (1988). They tried to teach patient HM the meaning of eight rare words that are not known by most people. Each word was presented once with a defining phrase as its definition. After this first presentation, a series of 60 blocks followed; in each block, HM was required to choose the appropriate definition, synonym, or sentence context for each word. Whereas normal controls mastered the task within a very few blocks, HM never mastered the task. He appeared to know the meaning of one of the words from the start, but he never mastered the meanings of any of the other words. A similar dismal picture emerges from reports of HM's performance on paired-associate learning tasks, though the testing has not been as extensive. On the McClelland et al. (1995) theory, we should have expected that with enough repetitions, brain regions outside the medial temporal areas would have acquired the new word meanings through the gradual accumulation of small changes, just as they acquire new cognitive skills. The first puzzle, therefore, is:

Puzzle #1: Given their impressive gains from repeated practice in some tasks, why has it proven impossible in some other tasks to teach amnesic subjects new information, even with massive repetition?

The second puzzle lies in the apparent dramatic reduction of the ability to acquire new phonological distinctions in adulthood. As before, the puzzle arises within the framework of a connectionist approach to processing and learning, in which cognitive change occurs through the adjustment of connection weights during the course of processing. Such connection adjustments are thought to underlie children's and adults' acquisition of many sorts of knowledge and different kinds of skills. These include their ability to differentiate contrasting phonemes in their native language, their semantic knowledge of the world, their acquisition of cognitive skills such as reading, and many other things. The puzzle arises from the fact that many kinds of learning that occur in childhood are still quite possible in adulthood, suggesting that connection strengths are still subject to change in many domains, but the ability to learn to distinguish some phonetic contrasts not used in a person's native language appears to be lost or at least much diminished in adulthood. One famous case in point is the distinction between the English phonemes /r/ and /l/, which is extremely difficult for many Japanese adults, whose native language lacks such a distinction. Such people experience great difficulty in both the production and the perceptual identification of these sounds,

and attempts at remediation have generally shown very gradual progress (details are presented below). Thus the second puzzle is:

Puzzle #2: Why is it that adults can still learn and adapt many skills, yet the ability to adapt the perception and production of speech appears to be diminish drastically in adulthood?

Others have offered answers to both of the puzzles I have described here, but each puzzle has generally been considered independently of the other and the answers have not been at all related. I will suggest that both of these puzzles can be resolved by considering the functional consequences of the mechanism of learning that was proposed by Hebb. I now turn to an examination of Hebb's proposal.

Hebb's Proposal for Learning

Hebb outlined in his proposal for learning in his book *The organization of behavior*:

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficacy, as one of the cells firing B, is increased. (Hebb, 1949, p. 62).

This proposal has been the subject of intense study by neurobiologists and has received a great deal of experimental support through research on the neurobiological phenomenon of long-term potentiation (see McNaughton, 1993 for a review). Briefly put, in a slice of brain that has been placed in a dish, one can study the effect of electrically-supplied pulses that emulate the firing of neuron A (and probably several others) on the activation (excitatory post-synaptic potential) and subsequent firing of neuron B. The magnitude of the EPSP is taken to reflect the synaptic efficacy. One can repeatedly excite A but, if strong depolarization of B does not occur, there is little or no change in the EPSP. But if conditions are arranged to pair the EPSP produced by A with a strong-enough depolarization of B, a long-lasting change in synaptic efficacy results (Barrionuevo & Brown, 1983). Interestingly, the EPSP from A must just barely precede the depolarization of B in order to have this effect (Markram & Sakmann, 1995); thus synaptic input from A that could be causal in firing B at the right time is strengthened, whereas synaptic input that comes too late to be causal is not.

Let us think for a minute about Hebb's rule, in the context of a situation in which some input (say the sound of a word) is presented. This input elicits activation of neurons throughout the auditory system and may elicit subsequent semantic or other associations. What Hebb's rule suggests is that the mechanisms of synaptic modification will tend to stamp in whatever pattern it was that the input turned out to elicit. Experiments on potentiation suggest that the stronger the elicited activation, the stronger the effect will be and the longer it will last. The result will be an increase in the probability and the efficiency of a subsequent, very similar input to produce the same activation. To the extent that the activation is appropriate and useful, effective acquisition and maintenance of desirable cognitive abilities will occur. But to the extent that the activation is inappropriate, Hebbian synaptic adjustment will tend to stamp in existing tendencies, and progress in acquiring the desired or appropriate response will not occur. I will argue in what follows that failures of learning

in amnesia and failures to acquire non-native speech contrasts in adult second-language learners might both arise from undesirable strengthening of inappropriate pre-existing activations.

Implications of Hebb's Proposal I: Learning in Amnesia

With the above background we can now consider why it may have proven difficult to teach amnesics arbitrary paired associates or meanings of unknown words. The answer may lie in part with the arbitrariness of the pairings to be learned, in part with the procedure used to teach them, and in part with the nature of the amnesic syndrome. In paired associate learning, the standard procedure is to select a set of arbitrary word-pairs, including, say LOCOMOTIVE-DISHTOWEL, and then to present the pairs, one at a time, for a single study trial. On subsequent learning trials, only the first or stimulus item is shown, and the subject is required to try to produce the second or response item. Generally, even normal subjects get few if any items correct on the first such trial, so after the subject guesses, a correction is provided. Now, the nature of the amnesic syndrome becomes crucial. As previously noted, the syndrome is thought to prevent the the rapid formation of arbitrary associations that bind the specific stimulus and response items in a pair together. In normals, we assume, such associations are formed during learning trials, and serve as the basis for the correct response. But the amnesic subject is impaired in the formation of such associations, so no retrieval can occur. Instead, the subject must rely on the small adjustments that were made to the strengths of connections outside the medial temporal region and on whatever pre-existing associations he or she may have to the stimulus. By assumption (motivated in McClelland et al., 1995), the small adjustments in the neocortex that occur during any given processing event are too small to create a sufficient association between arbitrarily paired items, so the subject is very unlikely to be able to produce the correct response. But in paired-associate learning, the subject is strongly enjoined to produce some response, which is very unlikely to be correct. But according to the Hebbian hypothesis, synaptic modification processes strengthen whatever neural activity occurs in response to the stimulus. The effect will be to strengthen the incorrect response that is elicited, rather than the desired one.

If there is any validity to this explanation, one would expect that amnesics would learn better if conditions are arranged to prevent them from making errors. In fact, such procedures have been used to induce learning of arbitrary associations in amnesics. One of these is called the *method of vanishing cues*, and essentially it involves providing enough of the correct answer to ensure that the patient never makes a mistake, then gradually diminishing the cues (Glisky, Schacter, & Tulving, 1986). Unfortunately, such training is usually carried out in practical contexts, and so it is difficult to run control groups to demonstrate unequivocally the importance of the prevention of error responses.

There are, however, two experiments that provide more specific tests of the effects of allowing patients to make non-target responses on their learning of target responses. In both cases, we see that if the training context is set up so that amnesics (or normal subjects, for that matter) will make non-target responses, subsequent performance on a test for desired or target responses will be impaired.

Baddeley and Wilson (1994) selected lists of single words, and studied amnesic's learning, in a test where they were required to generate the whole word from the first two letters. Two conditions were run. In the *errorful* condition, subjects were required to guess responses for each of the target words. For example, the experimenter might say, "I am thinking of a five-letter word

beginning with QU. Can you guess what the word might be?”. If the subject did not guess the pre-designated word (e.g. QUOTE) within four guesses or 25 seconds, the experimenter told the subject the word. If the correct response was the subject’s first response, an alternate word was substituted as the “correct” response (e.g. QUILL), so that at least one incorrect guess was elicited for each cue. After either guessing the word correctly (on the second or a later try) or after being told the correct word, the subject was asked to write down the correct word. A total of three trials of this type were run for each item, although the target remained fixed for the second and third trials, and correct responses at the beginning of these trials were allowed to stand. In the contrasting, *errorless* condition, no opportunity to guess was provided. Instead the subjects were told (for example) “I am thinking of a word beginning with QU. The word is QUOTE. Please write it down.” Again three trials of this type with each item were presented. It is important to note that in both conditions, the subject always wrote down the correct target word on each trial.

After this initial phase, the subjects, who were amnesics, elderly controls, or young controls, were given standard memory instructions. The experimenter said (for example), “Earlier I told you I was thinking of a word beginning with QU. What word was it?”. The striking finding (illustrated in Figure 1) was that amnesics in the errorful condition only obtained about 30% correct on the first trial, and they showed little improvement thereafter, whereas in the errorless condition they obtained 70% correct on the first trial and progressed relatively quickly to better and better performance. Both control groups showed similar effects, though of lesser magnitude.

Another experiment by Hayman, MacDonald, and Tulving (1993) produces a similar finding. In this case a single amnesic subject known as KC was pre-tested on a set of silly definitions for words. Two examples were “a talkative featherbrain” and “Marlon Brando’s mother”. The targets for these definitions are PARAKEET and the GODMOTHER. The set was then divided into two sets, one containing all those for which KC was able to provide a pre-existing, but incorrect answer (high interference items) and one containing only items for which he had no answer (low interference items). These two sets were then further subdivided into two halves, each of which was used in one of the two following conditions. In the *evaluate only* condition, KC was simply given a silly definition and the target and asked to evaluate how easily another person like himself would be able to grasp the connection between the definition and the word. In the other, *generate and evaluate* condition, KC was first asked to generate his own response to the silly definition, and only then was he presented with the target and the definition to perform the same evaluation. In a final test, KC was asked to generate a response for all items. The findings were particularly striking for the high interference items. In the evaluate only condition, KC gave 67% correct responses, and the probability of making a non-target response declined from a pretest value of 79% to 29% in the post-text. In the generate and evaluate condition, however, KC gave only 12% correct target responses after training, and during the post-test he made incorrect, non-target responses at the same high 79% rate as in the pretest.

Together, these experiments seem to show that amnesics are particularly prone to fail to learn target responses when training trials give them opportunities to generate pre-existing, non-target responses to the stimulus item. Preventing such responses leads to far better and more rapid learning. This finding is certainly consistent with the idea that the elicitation of a response strengthens it, whether that response is desirable or not, in accordance with Hebbian learning.

It may be worth coming back to the learning of arbitrary associations, since it can be noted that the material considered in Hayman et al. (1993) and Baddeley and Wilson (1994) is hardly arbitrary. Indeed, both experiments rely on materials where there is either a clear pre-existing

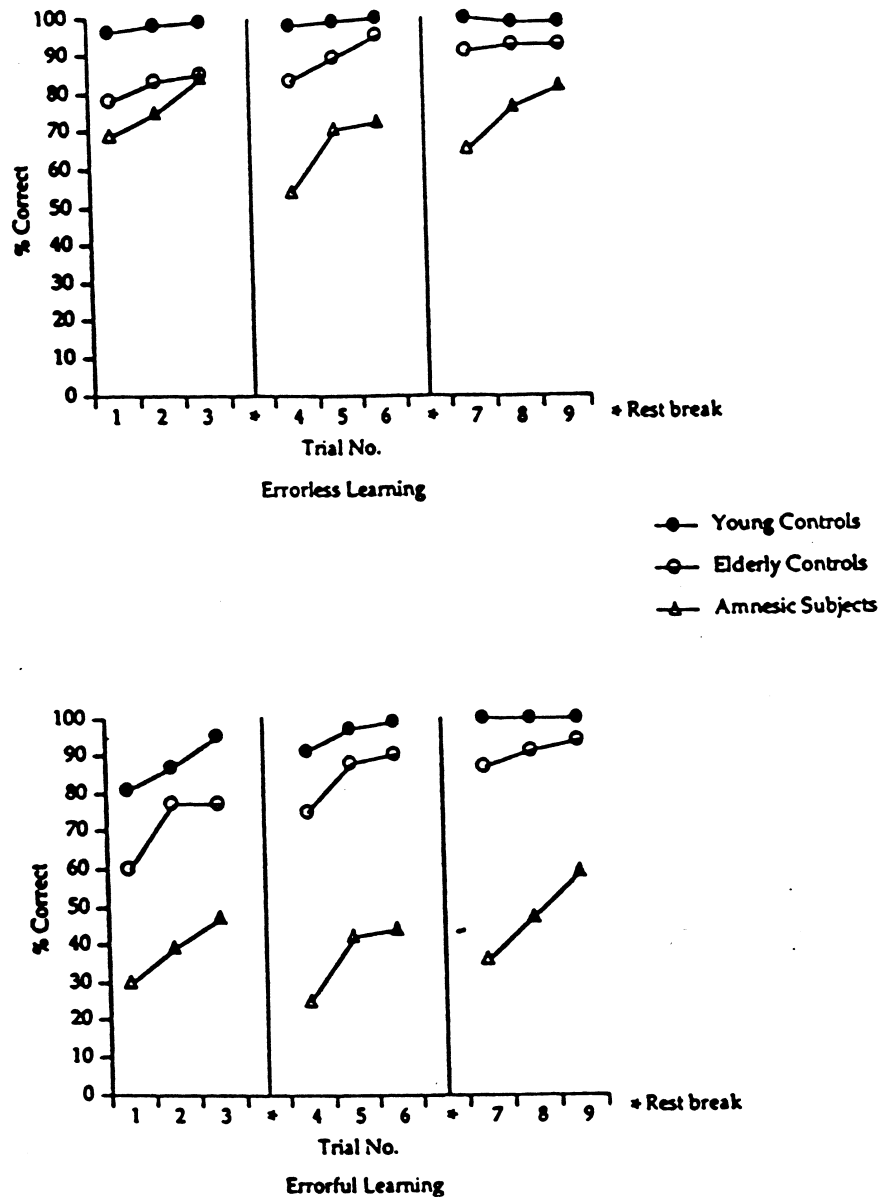


Figure 1: Percent correct cued recall of each of several test trials for words previously presented by the experimenter, under errorless and errorful learning conditions. From Baddeley and Wilson (1994). Reprinted with permission.

connection between cue and target (as in the first two letters of a familiar word and the rest of that word) or a possible way of making some meaningful connection, however silly, between the target and the cue, which makes the association far from completely arbitrary. We would expect amnesics to make more rapid progress when there is some pre-existing basis for association than they would when the associative relationship between cue and target is completely arbitrary. Our theory claims that the mechanisms of synaptic modification for synapses outside the medial temporal area are deliberately set up so that connection weights change only a little on each episode of processing. This gradual learning facilitates discovery of useful internal representations and avoids catastrophic interference (McClelland et al., 1995), but the downside is that the acquisition of truly arbitrary associations in these connections is necessarily very gradual. Thus, one would not expect errorless learning conditions to lead to rapid learning of arbitrary associates; these would still require many repetitions, even if inappropriate associative activations are prevented.

This discussion of amnesia brings out one important aspect of the role of the hippocampus in individuals with intact hippocampal systems. In the McClelland et al. (1995) model, the hippocampal system plays a very important role in allowing the contents of recent episodic memories to override whatever pre-existing associations may be present in connections outside the medial temporal area. This role seems particularly apparent in the control groups of the Baddeley and Wilson study, where there is only a very small decrement in performance in the errorful condition. This hints at a general point, which is that for Hebb's proposal to be at all feasible as a mechanism for learning, there must be a great deal of support for desired responses, including both supporting mechanisms within the brain and contextual support from the environment. Without this support, Hebbian systems would be far too susceptible to the undesirable effects of the self-reinforcing characteristic of Hebbian learning.

Implications of Hebb's Proposal II: Critical Periods in Learning Phonology?

We have seen how the mechanism of Hebbian learning might reinforce incorrect responses elicited in memory experiments, thus leading to failures to benefit from practice in experiments where such responses are routinely elicited. We will now consider how the same idea may help us understand why adults sometimes fail to learn to distinguish speech sounds that are not contrasted in their native language. To begin to see how this might arise, imagine an individual who has had years of experience carving up phonological space according to the phonological structure of her native language. Research on speech perception in infants suggests that when this person first came into the world, she would have been capable of distinguishing the phonemes characteristic of any natural human language. We also know that within a relatively short period of life, the baby would lose the ability to make most of the distinctions not made in her language environment. There are different theories of how this loss may arise. In our view, a useful framework is provided by Kuhl's (1991; Kuhl & Iverson, 1995) perceptual magnet theory, in which the phonetic prototypes of one's native language act like magnets or (in neural network terms) attractors, distorting perception of items in their vicinity to make them more similar to the prototype (See also Samuel, 1982). Suppose the individual we are considering had been born into a Japanese language environment, where there is only a single alveolar liquid phoneme instead of the two distinct phonemes /r/ and /l/. For this individual, a single perceptual magnet would arise in the perceptual space spanned by English /r/ and /l/. As the child grew older this magnet would grow stronger and stronger, so that different items within the vicinity of this magnet would come to be more and more strongly attracted to it.

Once this process had become quite strong, presentation of either an /r/ or an /l/ would, through the operation of the perceptual magnet effect, result in a pattern of neural activation corresponding to the single Japanese alveolar liquid. Let us suppose this person then moved to the United States, where there would be separate /r/ and /l/ phonemes. What would happen when this Japanese adult hears examples of these phonemes? Based on past experience, tokens of either phoneme will give rise to the alveolar liquid representation. Now the mechanisms of Hebbian synaptic adjustment will have their paradoxical, undesirable effect: Every time either an /r/ or an /l/ is presented, Hebbian synaptic modification will simply stamp in the tendency for the sound presented to activate this Japanese perceptual representation. The result will be that the phonological categories the person brought with her from Japan will simply be reinforced rather than eliminated by experience in the new environment.

A Computational model.

For these suggestions to be sufficiently explicit and concrete, and to show that they actually lead to the envisioned effects, it seemed clear that a computational model was needed. Adam Thomas and I have developed such a simulation model (McClelland & Thomas, 1998), and I will describe it briefly here. As in other modeling work I have been involved in, the approach has been to develop a model with the minimal structure necessary to capture the intuitive account with which we began. In my experience even the simplest models are complex enough, and one can learn a great deal from them. In addition to simplicity, we wanted a model that would begin to make some contact with domain-general aspects of the physiology of perceptual representations. These characteristics were already present in a model of Kohonen (1982; 1990) called the Self-Organizing Map, so we adopted the Kohonen model, adapting it slightly for our purposes. The architecture of the model is shown in Figure 2. It consists of two layers of 49 units, with the units in each layer arranged in a two-dimensional sheet. These layers are called the input layer and the representation layer. There are modifiable connections to each representation layer unit from each input layer unit; the connections to one of the representation units are illustrated in the Figure. Others have previously used a very similar model to capture perceptual magnet effects (Guenther & Gjaja, 1996), although we will not be considering these further here.

The model is thought of as experiencing patterns generated from a simple environment. There are two environments, one analogous to Japanese and one analogous to English, and these are shown on the bottom panel of the figure. In the English environment, inputs are generated from each of six prototypes. Four are called *corner prototypes* (labeled A in Figure 2), and they are thought of as analogous to background phonemes in English. The other two are called *overlapping prototypes* (B), and they are thought of as analogous to English /r/ and /l/. In the Japanese environment, inputs are generated from the same four corner prototypes, but in place of the two overlapping prototypes there is a single *central prototype* (C) analogous to the Japanese alveolar liquid.

The model works as follows. Inputs are generated by selecting one of the prototypes, perturbing it with a bit of noise, and then setting the activations of the input units to the resulting perturbed values. An example case from the left overlapping prototype is shown imposed on the input units in the diagram of the model. Based on the existing connection weights, this input pattern generates excitatory influences on all the units in the representation layer (note that activations and weights are both constrained to range from 0 to 1 in this model). The net input to each representation unit is

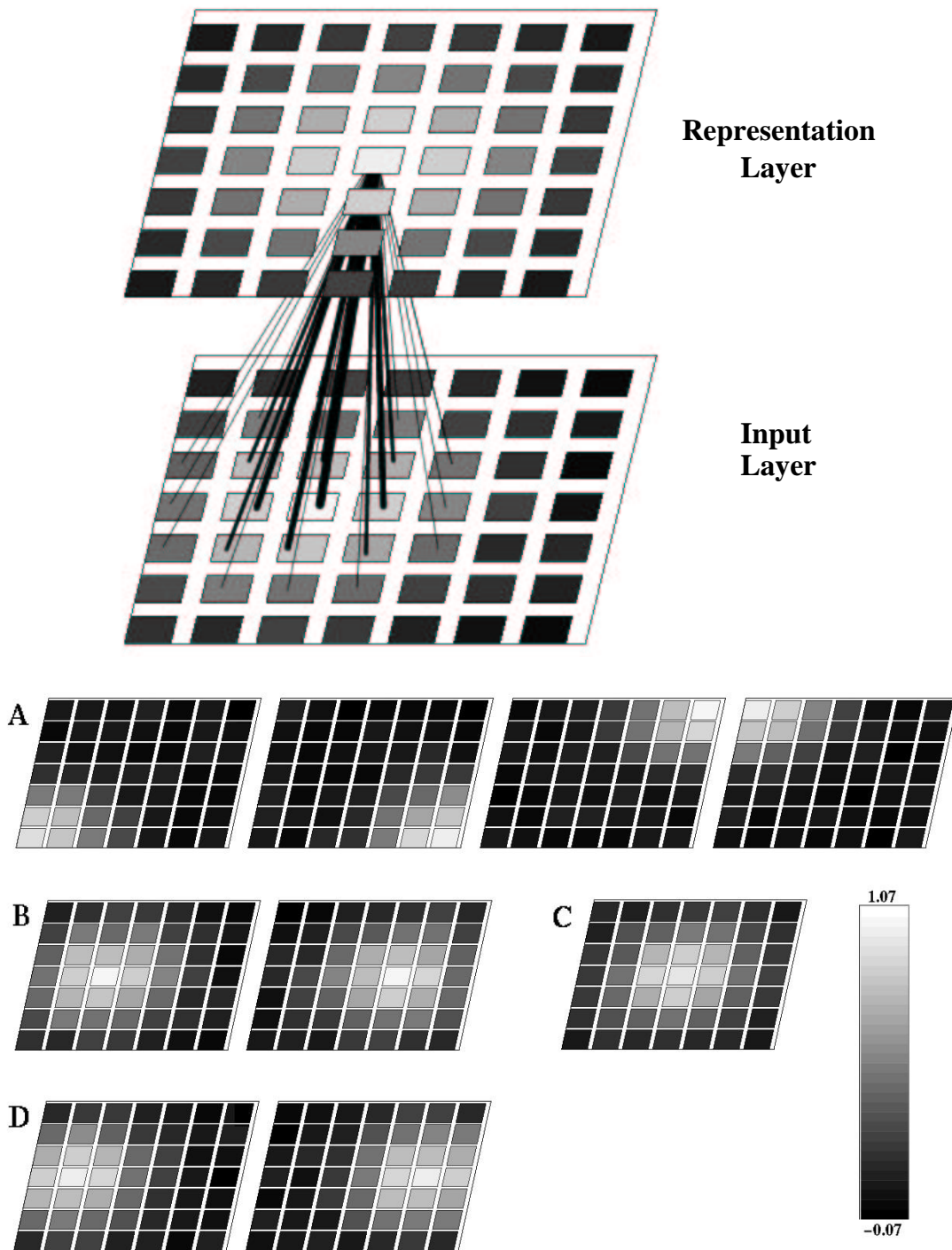


Figure 2: Top: The network used in the simulations, based on the self-organizing map model of (Kohonen, 1982; 1990). Bottom: The prototypes of the input patterns used in training and testing the network. (A) The four corner prototypes. (B) The two overlapping prototypes. (C) The single central prototype. (D) Exaggerated versions of the central prototypes used in remediation of the network.

calculated: this is the sum, over all of the unit's incoming connections, of the strength of the weight on the connection times the activation of the input unit at the other end of the connection ($net_r = \sum_i w_{ri} a_i$, where net_r is the net input to representation unit r , a_i is the activation of input unit i , and w_{ri} is the weight to representation unit r from input unit i). The unit with the largest net input is chosen as the winner, and its activation is set to 1. The activations of other units around the winner are set according to a Gaussian function of distance from the winner, so that near neighbors of the winner have reasonably strong activations and far neighbors are inactive. These aspects of the model are taken to reflect the influence of short-range excitatory and longer-range inhibitory interconnections among neurons in sensory cortex. Together with the connection weight adjustment rule about to be described (and variants of it), they have been used successfully to account for a number of aspects of neural organization, including the emergence and maintenance of topographic maps in sensory cortical areas as well as for the tendency for neurons representing similar higher order features to be located near each other in inferotemporal cortex (Wang, Tanaka, & Tanifuji, 1996).

In the model, once a representation has been assigned to an input, the connection adjustment rule is applied. The form of the rule that we use is often called the *competitive learning* as is also known as the *Oja rule*, based on its uses by Oja (1982) as in a principal-component analyzer. The rule is essentially Hebbian in nature, but with a built-in normalization factor. The rule is

$$\Delta w_{ri} = \varepsilon a_r (a_i - w_{ri}) .$$

In this rule, ε is the learning rate constant, and $a_r a_i$ is the Hebbian element, increasing strengths of connections to active representation units from active input units. The remaining subtractive element of the rule ($-a_r w_{ri}$) tends to cause weights from inactive input units to be reduced, and normalizes the weights to a representation unit so that $\sum_i w_{ri} = \sum_i a_i$. In our model, $\sum_i a_i$ is approximately constant (except for effects arising at the edges of the network), so that the sum of the weights coming to any representation unit tend to be fairly constant as well.

We can consider the learning rule at a slightly more abstract level and note that what it tends to do is to align the weights of the winning unit and its neighbors to match the current blob of activity on the input units. The rule can be seen as pulling the existing blob of weights toward the current input blob, and thus of implementing Hebbian learning at a system level by reinforcing the tendency of the current input to activate the same winning unit next time.

We simulated an analog of critical period effects in learning the phonological structure of English using this model, as follows. Each of 1000 different simulated learners was simulated with a different copy of the same network. Initial connection weights to each representation layer were set to random values with a loose topographic structure so that the initial weights to a particular representation unit tend to be strongest from corresponding points on the input layer and tend to fall off with distance. (This appears to characterize at least roughly the initial topographic biases in layer-to-layer projections in the brain.)

Our first baseline simulation illustrates what happened when networks were trained from the beginning with the "English" environment consisting of patterns generated from the four corner prototypes and patterns generated from the two overlapping prototypes. 1000 different networks, analogous to 1000 different children learning English from birth, were randomly initialized and then trained according to the following protocol. Training occurred in epochs. In each epoch one training example was generated from each of the six prototypes in random order. After the presentation of each pattern, the winner was chosen and the weights were then adjusted as described above. Periodically we tested the network, with learning turned off, and looked to see whether the network differentiated the two overlapping patterns, as English children differentiate the phonemes

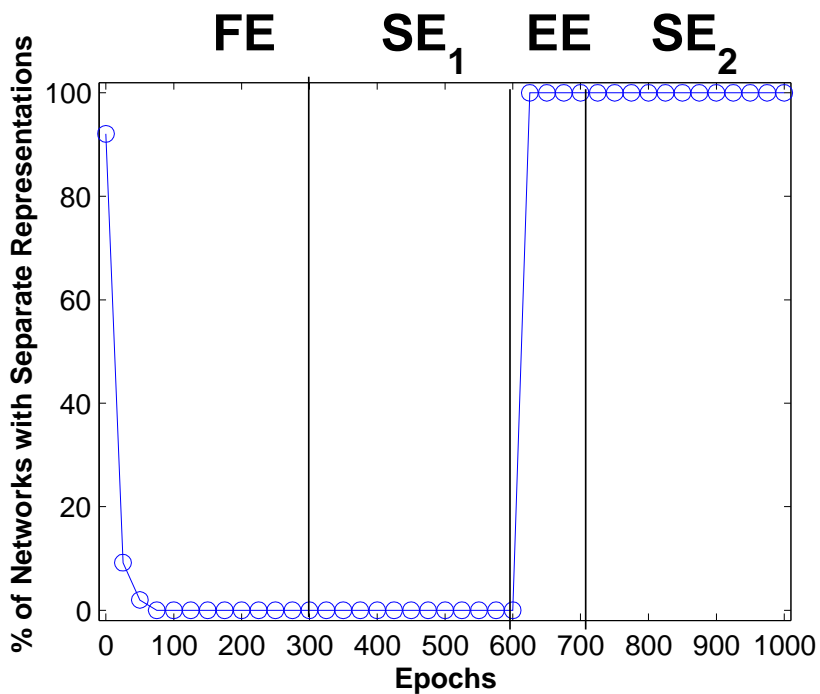
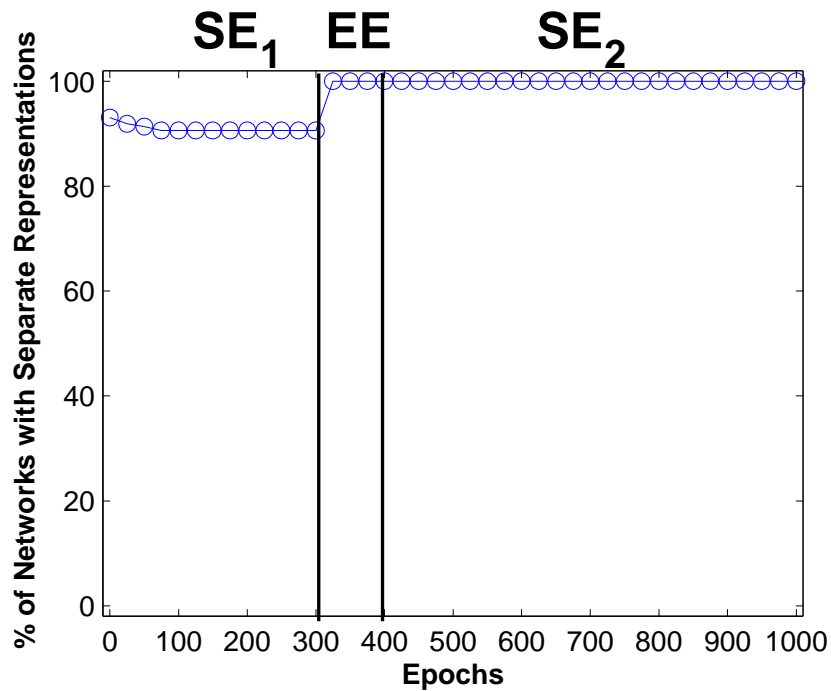


Figure 3: Simulation results from the model of McClelland and Thomas (1998). Vertical axis indicates the number of networks that assigned distinct representations to the two overlapping prototypes at each time point. (A) Results of training on the six-prototype “English” environment. Remediation training with exaggerated versions of the overlapping prototypes occurred between epochs 300 and 400. (B) Results of training on the six-prototype “English” environment (starting at epoch 300) after initial training on the five-prototype “Japanese” environment. Remediation training with exaggerated versions of the overlapping prototypes occurred between epochs 600 and 700.

/r/ and /l/. Differentiation was indexed by whether the two patterns activated different representation units.

The results of the first simulation are shown in the top panel of Figure 3. Right from the start, about 92% of the networks assigned distinct representations to the two overlapping prototypes, and this tendency continued throughout 300 epochs of training. The 8% of the networks that failed could have been manipulated up or down; we chose to leave the parameters as is, to invite consideration of possible bases of perceptual deficits among native speakers of the English language. Not apparent in the figure is the considerable tuning and structuring of the representations of the inputs that actually results from the synaptic modification process as this occurs during learning. The effects of this will be apparent, however, in the next simulation.

The second simulation considered the effects of initial exposure to the “Japanese” environment. 1000 networks were initialized as before, but this time initial training over 300 epochs was undertaken using the four corner prototypes and the single central prototype. Periodic testing using the overlapping prototypes with learning turned off revealed that over the 300 epochs, the networks learned to treat the two overlapping prototypes from the “English” environment (analogous to English /r/ and /l/) as the same: all the networks came to assign both overlapping prototypes to the same representation as shown in the bottom panel of the Figure. A case of a network with this tendency is actually shown in Figure 2: An example of the left overlapping prototype is shown activating the representation unit in the middle of the representation layer. This same unit is also activated in this network by presentations of the right overlapping prototype.

At this point we are finally ready to ask what happens to such a network when its language environment changes, as would happen in the case of a Japanese adult who moves to the United States. In the model the analog of this is a shift in the training environment from the “Japanese” (five-pattern) environment to the “English” (six-pattern) case. The result is shown in the next section of the bottom panel Figure 3. At epoch 300, the training environment was switched from the five-prototype, Japanese-like environment to the six-prototype, English-like environment. Yet it is apparent in the figure that the exposure to the six prototype environment does not result in a change in the number of categories represented by the network. Every one of the 1000 networks persists in maintaining a single representation, spanning both the left and right overlapping prototypes. This behavior continues indefinitely, and reflects the fact that whenever an example of either the left or the right prototype is presented, the same unit in the representation layer wins. The examples do jockey the weights back and forth a bit, but the jockeying averages out, and the network remains stuck with a single representation that spans the two overlapping prototypes.

In short, the network shows how a Hebbian learning mechanism can lead to a situation in which the network learns to treat two classes of inputs as the same and then diabolically maintains this tendency, even when faced with input that would at first have caused it to represent the classes separately. We will consider below how this aspect of the network’s behavior actually depends on additional aspects of the Kohonen network beyond the use of a Hebbian learning rule, but for now, we may consider whether the Hebbian perspective can suggest any steps we might take to overcome the network’s tendency to fuse the two overlapping categories.

Note that our approach suggests that the mechanisms of synaptic modification are still at work in the network; the only problem is that whenever an example from the right or the left prototype is presented, the same representation is activated, and the mechanisms of synaptic modification are actively maintaining the network’s tendency to map examples of both prototypes to the same representation. To fix the problem, what we need to do is find some way of having the inputs activate

different representations. In this context, consider what might happen if we used inputs that exaggerate the difference between the left and the right prototypes. If they are exaggerated enough, they will activate distinct representations. Once this happens, the mechanisms of synaptic modification will strengthen separate representations of the exaggerated items. We can then gradually reduce the difference between the inputs, making sure that we maintain the network's tendency to assign them distinct representations. Under these circumstances, the Hebbian synaptic modification process will tend to reinforce the tendency to treat the stimuli as distinct. Following this procedure we may eventually be able to move the inputs back to their original positions. We could even make them more similar to each other than they ordinarily would be in the natural situation, and still maintain separate representations.

Because of the relatively coarse grain of the model, the procedure we used in the simulations was even simpler than what I have just described. We simply added examples taken from exaggerated versions of the overlapping prototypes, shown in Figure 2D, to the six-prototype training environment, so that now there were eight stimuli in each epoch, one from each of the corner prototypes, one from each of the original overlapping prototypes, and one from each of the two exaggerated versions of the overlapping prototypes. This approach was very effective. Within 25 epochs, all networks had split the central representation into two distinct representations. When after 100 epochs the training environment reverted to the standard English-like six-prototype environment, all networks maintained separate representations of each of the two overlapping prototypes. The same intervention was also effective in remediating all of the networks that had been trained from the start with the English-like environment but had not separated the overlapping prototypes. In fact, in other simulations we have shown that various conditions analogous to possible causes of language impairment can lead the network to form a fused representation of the two overlapping prototypes; unless these conditions are extreme, the remediation training causes the representations to separate (McClelland, Thomas, McCandliss, & Fiez, 1999). The separation will be maintained even after the exaggerated examples are removed, except when the initial deficit is extremely severe. As one example of the possibilities we considered, one might suppose that in some cases of language impairment, crucial distinctions between certain phonemes are less salient than in most children, due perhaps to differences in processing rapid transitions (this might affect discrimination of stop consonants such as /b/ and /d/ more than liquids). A simple analog of this in our simulations involves increasing the overlap of the left and right overlapping prototypes, to the point where they almost always result in a single, fused representation. In this situation, remediation with exaggerated examples results in differentiation of the representations of examples of the overlapping prototypes, and unless the increase in overlap of these prototypes is extreme, the separation will be maintained after the exaggerated training examples are removed.

The remediation training we used in our network is similar to the procedure used by Merzenich, Tallal and their collaborators in their intervention studies with language impaired children (Merzenich, Jenkins, Johnson, Schreiner, Miller, & Tallal, 1996; Tallal, Miller, Bedi, Byma, Wang, Nagaraja, & others, 1996). Children received extensive exposure to speech with exaggerated (amplified and temporally extended) transitions, and also received ongoing exposure to normal speech from their parents and peers. The children also received extensive training with computer games in which they received initially exaggerated contrasts which were then gradually reduced as their ability to discriminate improved. Another, less extensive study performed sometime earlier (Alexander & Frost, 1982) found that only a few sessions of exposure to exaggerated contrasts could lead to improvement in discrimination of stop consonants. In both studies, controls who received no ex-

posure to exaggerated speech showed much less improvement on average.

Experimental test of the implications of Hebb's proposal.

The thoughts about the implications of Hebbian learning that gave rise to our model also give rise to hypotheses about the key processes at work in the remediation studies just discussed. With several colleagues, I have sought to explore this matter further by running additional remediation studies (McCandliss, Fiez, Protopapas, Conway, & McClelland, 1999). In these studies, we had two goals.

First, we wanted to demonstrate remediability in adults, to test predictions of loss-of-plasticity approaches against those of our plasticity-maintains-stability account. It has been shown in some domains (e.g., plasticity of the coordination of auditory and visual representation of external space in the barn owl, (Brainard & Knudsen, 1998)) that animals are less susceptible to re-mapping after puberty than they are earlier in life. These studies, coupled with age effects on degree of mastery of a non-native language (Johnson & Newport, 1989), have led some to suppose that synaptic modification may be shut off or at least greatly reduced after puberty in task-relevant brain regions. Evidence that the right intervention could effectively re-open the critical period might lead to some rethinking of those views. Of course it is quite possible that mechanisms like those we have described play an important role in maintenance of prior language habits, and that at the same time there is some more maturationally-based reduction of the degree of plasticity in adults. We do not intend to suggest that such maturationally-based effects do not occur, but only that, in addition to any such effects, the Hebbian mechanism we have described may be a significant contributing factor. If indeed the mechanisms of synaptic modification are working strongly against learning under naturalistic circumstances, that it may be possible to alter circumstances so that these mechanisms can be turned to advantage instead of disadvantage. The question becomes, can we find an intervention that would exploit a Hebbian learning process and lead to rapid progress learning to distinguish /r/ and /l/ in Japanese adults, indicating that indeed considerable plasticity still exists if only we can harness it?

There have been several studies that have demonstrated training effects on discrimination of /r/ and /l/ in Japanese adults. Strange and Dittmann (1984) produced noticeable changes after several days of training with synthetic speech stimuli from a continuum of spoken inputs ranging from "rock" to "lock". In an extensive series of studies, Pisoni and colleagues have shown that they can improve discrimination of /r/ and /l/ through the use of natural speech tokens (see Akahane-Yamada, Tohkura, Lively, Bradlow, & Pisoni, in press for a summary of this work). Their approach involves presenting subjects with a naturally spoken word (such as "rake" or "eagle") and then giving the subject a two-alternative forced choice between the item and a minimal contrast foil ("rake" vs "lake"; "eagle" vs "eager") with feedback. While there is progress in the Pisoni studies, it is fairly slow; for example in the most recent study, an improvement of about 20% required 45 hours of training (Akahane-Yamada et al., in press). Thus the existing evidence does not provide strong support for the idea that there is still a lot of underlying plasticity for learning phonology in adults. On the other hand, the Pisoni group's training procedure does not seem particularly conducive to revealing the existence of this plasticity, since it is quite possible that the conditions of the experiment are promoting self-maintenance of existing tendencies to hear /r/ and /l/ as the same, at least for some of their stimuli. Their data show that for a subset of the stimuli they used, initial discrimination ability was very close to chance. Our goal was to see whether, with the use of

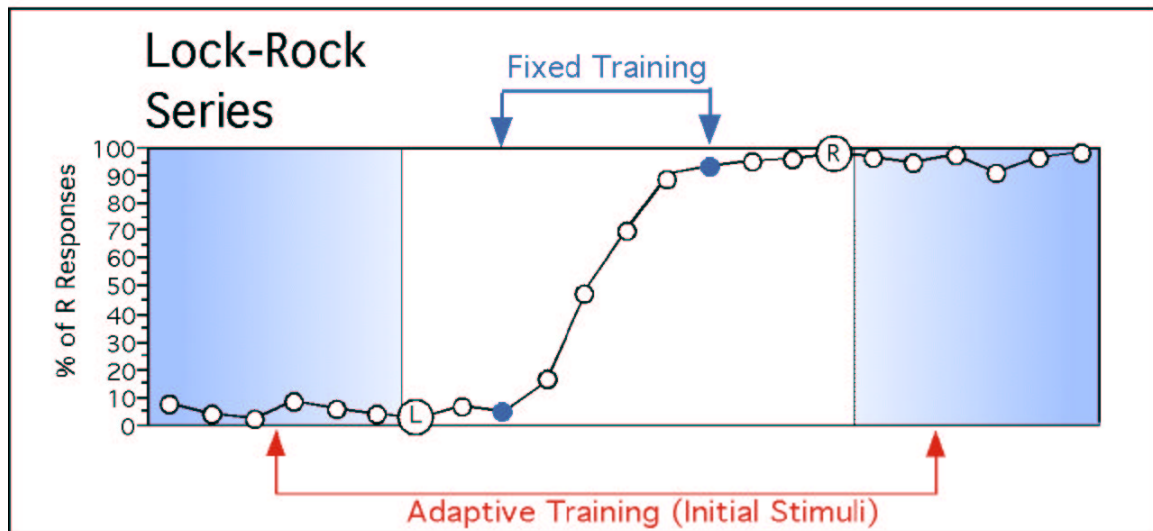


Figure 4: One of the two /r/-/l/ continua used in the training experiment with Japanese adults. Data indicating the phoneme identity assigned to these stimuli come from a group of University of Pittsburgh undergraduates. The stimuli used in the fixed training condition and the starting stimuli used for the adaptive training condition are marked. The two anchor stimuli which are resynthesized from natural tokens or “lock” and “rock” are circled and labeled with the letters L and R respectively.

exaggerated stimuli that subjects can initially discriminate, we could more rapidly induce learning of the /r/-/l/ contrast in Japanese adults.

Second, we wanted to test two aspects of the Hebbian approach to critical period effects in phoneme learning. One of these is that progress in learning can occur without feedback about correct performance. According to the Hebbian approach, as instantiated in our model, connections that map given inputs to the perceptual representations they produce are reinforced by Hebbian learning, even in the absence of feedback. The other aspect of the Hebbian approach is that success in learning to discriminate stimuli should depend critically on whether the stimuli used in training elicit different representations. As illustrated above in our simulations, the use of stimuli that are discriminable by native speakers of a language may not lead to differentiation of perceptual representations if they are not discriminable to the second-language learner. But the use of exaggerated stimuli that are discriminable by the second-language learner should lead to much more rapid progress in learning.

We therefore contrasted two training regimes: (See Figure 4). One regime, called *fixed* training, used difficult minimal pair stimuli, specifically chosen to be very hard for our Japanese adults, while yet being reliably identifiable as /r/ and /l/ by native English speakers. The other, called *adaptive* training, used an entire continuum of stimuli ranging from exaggerated versions of one member of the difficult minimal pair at one end of the continuum for exaggerated versions of the other member of the difficult pair at the other end. Each subject in this condition was started with exaggerated items, and whenever the subject made a mistake, the stimuli were made even more

exaggerated, until the outer limits of our continuum were reached. Whenever the subject correctly identified eight stimuli in succession, the discrimination was then made more difficult. According to the Hebbian approach and the model outlined above, we would expect the adaptive training regime to lead to learning. On the other hand we would expect that subjects in the fixed training condition would have great difficulty, given that they were not initially able to distinguish the stimuli used in that condition.

Two different continua were used in the study one based on “rock” and “lock” (as shown in the figure) and one based on “road” and “load”. For screening and to provide a pre-training baseline, each subject was pretested on both continua. During pre-testing and post-testing, we assessed each subject’s identification of the anchor stimuli, which were resynthesized from natural tokens of “rock” and “lock”, and of all of the stimuli intermediate between these two items. We also measured discrimination of several pairs of stimuli, including the fixed training stimuli, several pairs that were less separated than the fixed training stimuli, and several pairs that were more separated. Subjects were allowed to participate in the experiment if their discrimination of the fixed training stimuli was below 70% correct on one of the two continua, which was then used for training in the experiment. After three days of 480 training trials, the same two tests that had been used during pre-training were re-administered. Eight subjects were run in the adaptive training condition, and eight in the fixed condition. In addition, we ran eight control subjects who received the pre- and post-tests with no training, only a three-day gap between the two administration of these tests.

As predicted, all of the subjects in the adaptive condition made considerable gains in their identification of /r/ and /l/ stimuli as a result of their participation in the experiment, as indexed by the results of a logistic regression analysis of their performance in identifying the anchor stimuli and the stimuli interpolated between them (McCandliss et al., 1999). Prior to training, these subjects showed relatively flat identification functions, but after training all subjects showed an increase in slope of the identification function, indicating the acquisition of the ability to assign the correct identity to the anchor stimuli and neighboring interpolated stimuli with a rapid transition in their labeling functions at some point between the two anchors. Results in the fixed condition were very different. Some subjects did appear to show improvements in their identification performance over training, and there was a slight overall improvement in the slope of identification functions, but it appeared that even this slight gain could be attributed to the effects of the pre-test, since a gain of the same size was also obtained for the control subjects who received only the pre- and post-tests, with no intervening training. Thus, as expected on the Hebbian account, the use of exaggerated stimuli with subsequent adaptive modification led to rapid progress, but the use of difficult stimuli with no adaptive modification produced little or no benefit.

The results of the experiment are promising in suggesting that there may indeed be a great deal of residual plasticity in the adult brain that may allow the acquisition of non-native speech discriminations. However, it is important to note that in our study, we have not completely reorganized the perceptual categories of our Japanese subjects. On the contrary, it appears that what they have learned is quite specific, in that the benefits of training on one of our two continua did not transfer to the other continuum. As others have found (Akahane-Yamada et al., in press), training across a wide range of tokens appears to be necessary to produce a generalizable improvement that transfers to other speakers and other items.

Discussion.

Our experiment has succeeded in demonstrating that rapid learning of the /r/-/l/ distinction is possible in Japanese adults, even without feedback, if stimuli they can discriminate are used in training. We also showed that learning will not proceed if stimuli they cannot reliably discriminate are used, even when these stimuli are reliably identifiable by native speakers of English. These findings are consistent with the Hebbian proposal and the account we have offered with it of the basis for critical period effects in learning phonological contrasts. However, there are several additional points that have emerged from our research that indicate to me and my collaborators that the Hebbian proposal in its simplest form is at best only a partial guide to the conditions under which there may be successes or failures in learning.

For one thing, Thomas and I have observed in simulations that our Kohonen network will eventually learn to separate two overlapping inputs that it has previously come to map to the same representation, unless additional inputs that compete for space in the representation layer are interleaved with presentations of the two overlapping inputs. That is, we train networks that initially learned on the Japanese environment in a training situation where they are exposed only to exemplars of the overlapping prototypes, the network will eventually come to assign distinct representations to the two inputs. Initially we considered this to be a characteristic of the Kohonen network that might be unrealistic, but our experience in pilot studies has begun to suggest to us that this property of the networks may not be so unrealistic after all. Based on this feature of our networks we have continued some of the subjects in the fixed training condition for several additional days beyond the three training sessions used in the experiment reported above, and we have found that several of these subjects eventually did learn the /r/-/l/ discrimination. Thus, it appears that an important factor in the maintenance of perceptual representations is competition from other perceptual representations that are maintained through exposure to language inputs. As a corollary of this observation, the fact that in our experiments we focus training on just a single contrast may be part of the reason for the rapid success of our adaptive training regime. We intend to test this point explicitly in further investigations.

Another observation that has emerged from our studies to date may appear to contradict a fundamental aspect of the Hebbian proposal. This is the finding that learning to discriminate /r/ and /l/ appears to be facilitated by the use of feedback. We have found in a small study with eight subjects (four receiving adaptive training and four receiving fixed training) that with feedback, progress is possible both for the adaptive subjects and for subjects receiving fixed training. Clearly, this contradicts a narrow version of the Hebbian proposal, in which no role is given for outcome information in learning. However, it may not be inconsistent with other versions, in which Hebbian synaptic modifications are by feedback signals. One algorithm called the reinforcement learning algorithm (Barto, 1992) modulates the degree of Hebbian learning with outcome information. This algorithm has been shown to be quite effective in training networks to solve problems thought to require a more explicit error-correcting algorithm, and there is circuitry in the brain that may well allow the broadcast of reinforcement information to many of the synapses that may participate in perceptual learning. Thus, it is quite possible that Hebb's proposal was partially correct, but that it should be expanded to allow for modulation by outcome information, and possibly by other factors such as emotional state (McGaugh, 1989).

Conclusion

In this chapter, I have considered Hebb's proposal for learning in the brain, and I have suggested that his proposal may provide a partial guide to understanding some of the circumstances under which experience fails to lead to improvements in performance. The proposal provides a common account of failures of learning in amnesics and in normal second language learners, and has been incorporated in a simple simulation model that captures some aspects of critical period effects in the acquisition of contrasts between phonemes not distinguished in one's own native language. It also leads to testable predictions about what sorts of training regimes might lead to success or failure in teaching second language learners speech contrasts that are otherwise quite difficult for them to learn. That said, some additional observations that have arisen both from our modeling work and our empirical investigations suggest that Hebb's proposal for learning may turn out only to be a partial guide to the conditions under which humans and other organisms can improve their performance as a result of experience. There are many reasons to suppose that a complete account of how learning occurs in the brain will go beyond Hebb's initial proposal, and many reasons to suppose that additional insights into the successes and failures of learning at a behavioral or functional level will emerge as this more complete account is developed.

References

- Akahane-Yamada, R. A., Tohkura, Y., Lively, S. E., Bradlow, A., & Pisoni, D. (in press). Effects of extended training on English /r/ and /l/ identification by native speakers of Japanese. *Perception & Psychophysics*.
- Alexander, D. W., & Frost, B. P. (1982). Decelerated synthesized speech as a means of shaping speed of auditory processing of children with delayed language. *Perceptual and Motor Skills*, 55, 783–792.
- Baddeley, A., & Wilson, B. A. (1994). When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia*, 32, 53–68.
- Barrionuevo, G., & Brown, T. H. (1983). Associative long-term synaptic potentiation in hippocampal slices. *Proceedings of the National Academy of Science, USA*, 80, 7347–7351.
- Barto, A. G. (1992). Reinforcement learning and adaptive critic methods. In D. A. White, & D. A. Sofge (Eds.), *Handbook of intelligent control: Neural, fuzzy, and adaptive approaches* (pp. 469–491). New York: Van Nostrand Reinhold.
- Brainard, M. S., & Knudsen, E. I. (1998). Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *Journal of Neuroscience*, 18, 3929–3942.
- Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, 6, 205–254.
- Gabrieli, J. D. E., Cohen, N. J., & Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and Cognition*, 7, 157–177.
- Glisky, E. L., Schacter, D. L., & Tulving, E. (1986). Learning and retention of computer-related vocabulary in memory-impaired patients: Method of vanishing cues. *Journal of Clinical and Experimental Neuropsychology*, 8, 292–312.
- Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, 100, 1111–1121.

- Hayman, C. A. G., MacDonald, C. A., & Tulving, E. (1993). The role of repetition and associative interference in new semantic learning in amnesia: A case experiment. *Journal of Cognitive Neuroscience*, 5, 375–389.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.
- Hinton, G. E., & McClelland, J. L. (1988). Learning representations by recirculation. In D. Z. Anderson (Ed.), *Neural information processing systems* (pp. 358–366). New York: American Institute of Physics.
- Johnson, J., & Newport, E. (1989). Critical period effects in second-language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78, 1464–1480.
- Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93–103.
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Baltimore: York Press.
- Markram, H., & Sakmann, B. (1995). Action potentials propagating back into dendrites triggers changes in efficacy of single-axon synapses between layer V pyramidal neurons. *Society for Neuroscience Abstracts*, 21, 2007.
- Mazzoni, P., Andersen, R. A., & Jordan, M. I. (1991). A more biologically plausible learning rule for neural networks. *Proceedings of the National Academy of Sciences USA*, 88, 4433–4437.
- McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (1999). *Success and failure in teaching the [r]-[l] contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception*. Submitted for publication.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375–407.
- McClelland, J. L., & Thomas, A. (1998). *Dynamic stability and adaptive intervention: Consequences of Hebbian learning?* Unpublished manuscript, Center for The Neural Basis of Cognition, 115 Mellon Institute, Pittsburgh, PA.
- McClelland, J. L., Thomas, A., McCandliss, B. D., & Fiez, J. A. (1999). Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. In J. Reggia, E. Ruppin, & D. Glanzman (Eds.), *Brain, behavioral, and cognitive disorders: The neurocomputational perspective* (pp. 75–80). Oxford, England: Elsevier.
- McGaugh, J. L. (1989). Involvement of hormonal and neuromodulatory systems in the regulation of memory storage. *Annual Review of Neuroscience*, 12, 255–287.
- McNaughton, B. L. (1993). The mechanism of expression of long-term enhancement of hip-

- pocampal synapses: Current issues and theoretical implications. *Annual Review of Physiology*, 55, 375–396.
- Merzenich, M. M., Jenkins, W. M., Johnson, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science*, 271, 77–81.
- Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.
- Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15, 267–273.
- O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, 8, 895–938.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986a). 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 (pp. 318–362). Cambridge, MA: MIT Press.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986b). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I: Foundations & Volume II: Psychological and biological models*. Cambridge, MA: MIT Press.
- Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, 307–314.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195–231.
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36, 131–145.
- Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagaraja, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, 271, 81–84.
- Wang, G., Tanaka, K., & Tanifuji, M. (1996). Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*, 272, 1665–1668.
- Zipser, D., & Andersen, R. A. (1988). A back propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331, 679–684.