JOURNAL OF MEMORY AND LANGUAGE 27, 213–234 (1988)

Some Criticisms of Connectionist Models of Human Performance

DOMINIC W. MASSARO

Program in Experimental Psychology, University of California, Santa Cruz

Connectionism offers a challenge to current information-processing descriptions of linguistic performance. Upon examination, however, models with the connectionist framework are found to be wrong in important respects or are too powerful to be meaningful. The following observations support these claims. The assumption of interactive activation (i.e., two-way connections between units) of specific connectionist models is shown to be both unnecessary and inconsistent with empirical results. Connectionist models with hidden units are demonstrated to be too powerful; they can simulate different types of results that are generated by different process models. Given the power of connectionist models with hidden units, they can describe results with unrealistic assumptions about the psychophysical relationships that are functional in the task. Connectionist models with hidden units are limited in theoretical value without postulating something like sequential stages of processing in which some categorization occurs before response selection. Notwithstanding these limitations, it is noted that other important properties of connectionism are to be found in existing process models of pattern recognition.

The present issue of this journal is devoted to research carried out within a supposedly new paradigm of psychological functioning. This issue is only one source of information among many that document the impact of what I refer to as connectionism. It is difficult for me to assess whether our science has previously witnessed such an apparent revolution. I see it as a revolution not only because so many investigators are adopting its program, but also because the question whether there is any worthwhile alternative has been proposed more than once. Perhaps Watson's (1913) call to behaviorism and Neisser's (1967) book had similar impact, but sheer numbers appear to be on the side of connectionism. My tack is criticize this valuable movement to reveal some fundamental limitations in the approach. Only by correcting these shortcomings will connectionism have the potential for a positive contribution to empirical and theoretical psychology. The methods and goals of psychological inquiry have not changed and our task remains as difficult as ever.

Before departing on this adventure, it is important to acknowledge that not only are there differences in our concept of connectionism, but also that the concept itself must be fuzzily defined. Some view connectionism as a unique enterprise; others view it as spanning a continuum from models of the interactive activation variety to those used in backpropagation. Even others believe that interactive activation models are prototypical information-processing models, whereas backpropagation are prototypically connectionist. Like Wittgenstein's concept of games, there is no critical defining (necessary or sufficient) property of connectionism. We are limited to instances of connectionism and specific connectionist assumptions that can be eval-

The research reported in this paper and the writing of the paper were supported, in part, by grants from the National Science Foundation (BNS-83-15192) and the graduate division of the University of California, Santa Cruz. Patricia Carpenter, Michael Cohen, William Estes, Ray Gibbs, and Alan Kawamoto provided helpful feedback on an earlier draft of the paper. Reprints may be obtained from the author at Program in Experimental Pyschology, University of Santa Cruz, Clark Kerr Hall, Santa Cruz, CA 95064.

0749-596X/88 $3.00

uated. Thus, my criticism will address specific connectionist models and assumptions. Advocates of connectionism might, therefore, claim that the criticisms do not address the validity of the connectionist approach. However, the specific models and assumptions appear to be prototypically connectionist, and their demise offers a strong challenge to the validity of the current connectionist paradigm.

Testing the Interactive Activation Assumption

Central to the framework of connectionism is the assumption of high interconnectivity among the processing units. This assumption contrasts with the relative modularity assumed by some information-processing models (Palmer & Kimchi, 1986) and the extreme modularity assumed by some other researchers (Fodor, 1983). Connectionist models are constructed of units at different levels. All knowledge is contained in the connections among the units and the operations that map the input into the output. The units interact with one another via connections among the units.

Interactive Activation Model of Word Perception

The interactive activation model of word perception was designed to account for context effects in word perception (McClelland & Rumelhart, 1981) and was extended to account for other phenomena (Rumelhart & McClelland, 1982). The model postulates three levels of units: features, letters, and words. Features activate letters that are consistent with the features and inhibit letters that are inconsistent; letters activate consistent words and inhibit inconsistent words; and most importantly, words activate consistent letters. It has been known since the time of Cattell that a letter is more accurately recognized in the context of a word than in context of random letters. Interactive activation explains this word advantage in terms of interactive facilitation from the word level to the letter level. What is important for our purposes is that top–down activation from the word to letter level is necessary for explaining the word advantage. Thus, interactive activation might be considered to be the backbone of the model, which is why the question of the necessity of interactive activation seems fundamental. Rather than attempting direct tests of this assumption, however, previous research has been primarily limited to whether a model with interactive activation could account for various phenomena.

In 1979, I examined whether there was evidence for a two-way interaction between the letter and word level in visual word perception. To anticipate the results, there was no evidence for two-way interactive activation between letter and word levels (Massaro, 1979). The experiment factorially combined six levels of letter information with four levels of orthographic context. The letter information was manipulated by varying how much a test letter looked like a *c* or an *e* by extending the horizontal bar different amounts from right to left. To the extent that the bar is long, there is good visual information for an *e* and poor visual information for a *c*. There were six different test letters, varying from a clear *c* to a clear *e*.

Orthographic context was manipulated by varying the orthographic context to either favor one or the other letter or to be neutral. Consider the letter presented as the first letter in the context *-oin*. Only *c* is orthographically admissible in this context since these three consecutive vowels *eoi* violate English orthography. Only *e* is admissible in the context *-dit* since the initial cluster *cd* is an inadmissible English pattern. Given these constraints, the context *-oin* favors *c*, whereas the context *-dit* favors *e*. The contexts *-tsa* and *-ast* can be considered to favor neither *e* or *c*. The first remains an inadmissible context whether *e* or *c* is present, and the second is orthographically admissible for both *e* and *c*. Appropriate contexts were constructed so that the test letter could be presented at each of the four letter positions in each of the four types of context.

The test string was displayed for 30 ms followed by a blank interstimulus interval that lasted between 5 and 210 ms. This interval was followed by a 30-ms masking stimulus composed of random letter features. The subject was asked to indicate whether an *e* or *c* was present in the test display. Subjects were instructed to make the best choice on the basis of what they saw.

Figure 1 gives the observed interaction of bar length and orthographic context across the four masking intervals. The probability of an *e* response increased with increases in the bar length of the critical letter. The results also show a gradual resolution of the critical letter with increases in processing time before onset of the mask: the curves across bar length are steeper with longer masking intervals. An *e* identification was more probable for the context in which *e* but not *c* was orthographically admissible than for the context in which *c* but not *e* was admissible. The two neutral contexts were intermediate and did not differ from one another. This context effect was larger at the more ambiguous levels of the bar length of the test letter. Context also had a larger impact on identification of the test letter at the two extremes of the letter continuum to the extent the masking interval was short. These two results are consistent with the general finding that the contribution of context is larger with ambiguous than unambiguous bottom–up sources of information. The results were described in terms of the integration of context and letter information as two *independent* sources of information. The quantitative results were predicted by a fuzzy logical model of perception (FLMP) that had been successfully applied in other domains, such as speech perception and sentence interpretation (Oden, 1978; Oden & Massaro, 1978). What is important for our purposes is that the FLMP assumes that the top–down influence from the word

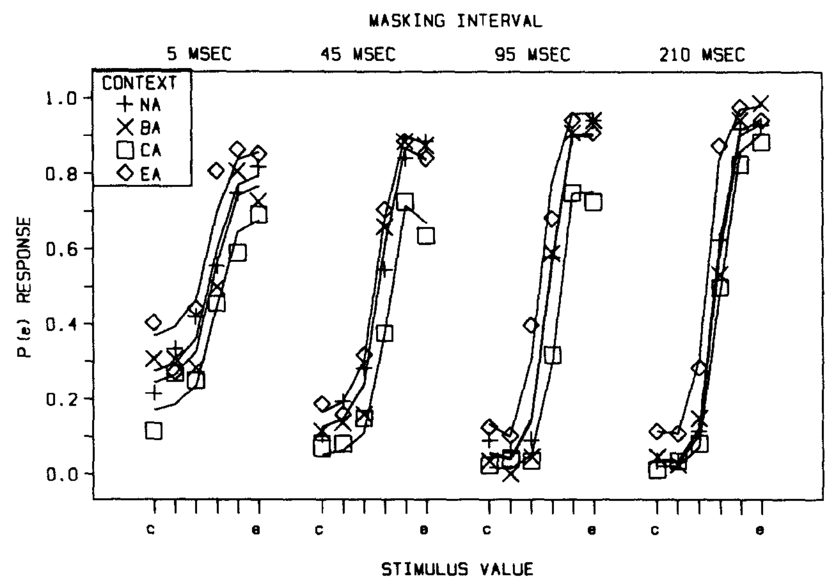

FIG. 1. Observed (points) and predicted (lines) probability of an *e* identification response as a function of the bar length of the test letter (stimulus value), the orthographic context, and the masking interval before the onset of the mask. The context NA = neither *e* nor *c* admissible, BA = both admissible, CA = *c* admissible, and EA = *e* admissible (results from Massaro, 1979). The predictions are given by the FLMP.

level does not modify the letter-level information.

The interactive activation model also predicts an effect of orthographic context, by assuming interactive activation between the units at the word and letter level. Presentation of *-dit* supposedly activates the word *edit,* and this activation feeds down from the word and activates its letters. Although the test letter in the display might have been somewhat ambiguous, the context can decrease the letter's ambiguity because *-dit* results in activation of the letter *e* at the letter level. A nonword context (*-tsa*) would be expected to produce very little activation from the word level to the letter level. In this case, the nonword context would not contribute to the discriminability between two adjacent letters along the letter continuum. The orthographic context (*-ast*) supporting both *e* and *c* would produce activation from the word level for both *e* and *c* at the letter level. This interpretation of interactive activation leads to the prediction that the orthographic context should have large effects on the discriminability of adjacent letters along the letter continuum.

Taking a signal detection perspective, we can describe the information being transmitted by the test letter. One measure of information is the degree to which the perceiver discriminates two adjacent letters along the test letter continuum (Massaro, 1979). The percentage of classification judgments of two adjacent letters along the letter continuum can be transformed into a measure of discriminability called, d', between the adjacent letters. The important assumption underlying this analysis is that this d' measure represents the relative activation of the *e* and *c* letter units in the interactive activation model. Using this measure, it is possible to test for d' differences independently of the actual effect of context on the overall likelihood of responding *e* or *c*.

The left panel of Fig. 2 gives the observed cumulative d' values across the stimulus continuum as a function of the four orthographic contexts. The d' values are different for the four contexts. There was no effect of context on the discriminability between adjacent letters along the letter continuum. Within the perspective of this analysis, the results indicate that the bottom–up activation of the letter was *not* modified by orthographic context, contrary to the expectation from interactive activation. Interactive activation predicts that the word context modifies the discriminability between adjacent letters along the continuum, relative to the nonword context. Consider processing of the test letter in terms of the interactive activation model. If an *e* is presented, bottom–up activation activates words consistent with *e* and inhibits inconsistent words. When *e* is presented in *-dit,* top–down activation from the word *edit* activates *e*. When *e* is presented in the context *-sta* there is very little top–down activation from the word to letter level. Thus, there is a large difference in the relative activation of the *e* and *c* letter nodes in the word than nonword context. The relative activation of *e* and *c* at the letter level determines the discriminability between the two test letters. Given that the relative activation of *e* and *c* differs for the word and nonword contexts, then their discriminability should also differ. The d' for the word context should be greater than the d' for the nonword context. Given that there were no differences in discriminability as a function of orthographic context, there is no evidence for interactive activation.

To illustrate that the data analysis was sensitive to variables that influence letter discriminability, the d' analysis was repeated as a function of masking interval. The right panel of Fig. 2 shows that discriminability of the letters along the letter continuum increases dramatically with increases in the processing time before the onset of the masking stimulus. Thus, the analysis does show changes in letter discriminability with masking interval, but

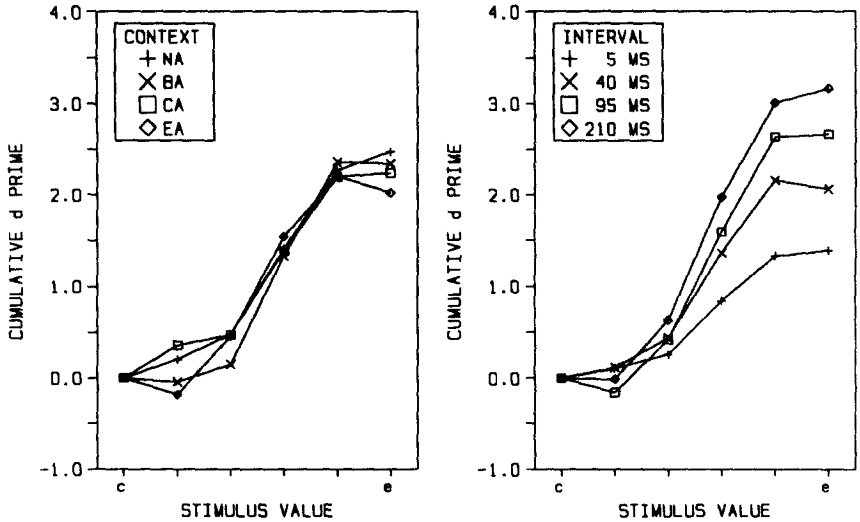

FIG. 2. Left: Cumulative d' values for the *e–c* continuum, as a function of the orthographic context. The context NA = neither *e* nor *c* admissible, BA = both admissible, CA = *c* admissible, and EA = *e* admissible (results from Massaro, 1979). Right: Cumulative d' values for the *e–c* continuum, as a function of the blank interval between the test word and the masking stimulus.

none with orthographic context. Hence, the analysis cannot simply be considered insensitive. Identical results were found in a replication using an *n* to *h* continuum and different orthographic contexts (Massaro, 1979). If this analysis is valid, it provides a straightforward disconfirmation of the central assumption of the interactive activation model of written word perception. It illustrates for me that demonstrating the adequacy of a model is not as productive as testing its fundamental assumptions within the spirit of falsification.

Interactive Activation Model of Speech Perception

The TRACE model of speech perception (McClelland & Elman, 1986) is a connectionist model in which information processing occurs through excitatory and inhibitory interactions among a large number of simple processing units. Three levels or sizes of units are used in TRACE: feature, phoneme, and word. Features activate phonemes which activate words, and activation of some units at a particular level inhibits other units at the same level. TRACE maintains the important assumption of interactive activation that activation of higher-order units activates their lower-order units; for example, activation of the /b/ phoneme activates the features that are consistent with that phoneme. For several phenomena, the predictions of TRACE are quantitatively similar to those of the FLMP. However, the interactive-activation assumption of top–down influence causes the two models to make very different predictions about categorical perception. Strictly speaking, perception is said to be categorical if the subject can discriminate stimuli between, but not within, speech categories. As described in Massaro (1987b, Chapter 4, Section 8), categorization in the FLMP occurs as a consequence of pattern classification and not at the featural evaluation or integration operations. As in the previous explanation of word perception, higher-order levels influence the decision without modifying lower levels. The TRACE model produces categorical-like behavior at the sensory (featural) level rather than at simply the decision stage. In this model, a stimulus pattern is presented and activation of the corresponding features sends more excitation to some phoneme units than to others. Given the assumption of feedback from the phoneme to the feature level, the activation of a particular phoneme feeds down and activates the features corresponding to that phoneme (McClelland & Elman, 1986, p. 47). This effect of feedback produces enhanced sensitivity around a category boundary, exactly as predicted by categorical perception.

The issue of categorical perception has been almost synonymous with speech perception research, and cannot be done justice here. However, we have learned something about where in the processing sequence categorization arises. There are several recipes for obtaining "categorical perception" results, but none of these can be taken as convincing evidence for interactive activation (Hary & Massaro, 1982; Massaro, 1987b). One recipe is to use a stimulus continuum that is irregular in terms of the psychophysical relationship between physical changes along the stimulus continuum and sensory/perceptual changes. One example is a /ba/–/da/ continuum in which /ba/ has rising formant transitions and /da/ has falling transitions. Discrimination of adjacent stimuli along the continuum most likely would be irregular because some pairs would be more discriminable than others. Specifically, we might expect enhanced discrimination between a stimulus with falling transitions and a stimulus with rising transitions. The reason is that the direction of change provides additional psychophysical information relative to the case in which two adjacent stimuli differ only in the absolute values of the formants, and not in the direction of change. What is important for TRACE is that enhanced discrimination at the point along the continuum having a change in direction of the transitions in no way depends on having interconnected feature and phoneme levels. The enhanced discrimination is also found for monkeys, who do not have these levels (Kuhl & Padden, 1983).

The second recipe for categorical perception is to have discrimination performance depend on an abstract context-sensitive code rather than a high-quality sensory-trace code. Thus, discrimination tests that make great demands on auditory memory tend to produce categorical perception results. Similarly, different segments produce different degrees of categorical perception results because their auditory representations differ in quality. Steady-state vowels produce a high-quality auditory memory and, therefore, do not give categorical results. Stop consonants, on the other hand, have poor-quality memories and give categorical results. The differential contribution of abstract and auditory memory is an important factor in producing categorical results. As in the recipe involving irregularities along the stimulus continuum, this recipe in no way depends on the interactive activation assumed by TRACE. The conclusion being advanced here is that these two recipes are the only contributions to categorical perception results.

In fact, categorical perception plays little or no role in speech perception. Language understanding requires categorical partition but this does not mean that categorical perception leads to this partitioning (Massaro, 1987a). Although the perceiver must determine if the speaker was referring to a ball or a doll, for example, the perceptual system could have continuous information representing the degree to which each alternative is supported. In fact, there is a growing body of research demonstrating that perceivers do, in fact, have continuous information (see Repp, 1984, for a review). Previous research had failed to contrast the predictions of categorical perception with those of a model of continuous perception. Given this limitation, our research carried out direct quantitative tests between categorical and continuous models of perception (Massaro, 1987b). In one series of experiments, subjects were asked to classify speech events that independently varied along two dimensions. The identification results were consistent with the assumption of continuous information along each of the two dimensions. A model based on categorical information along each dimension gives a very poor description of the identification judgments. In a second series of experiments, subjects were asked to make repeated ratings of the degree to which a stimulus represents a speech category. The distribution of the rating judgments to a given stimulus was more adequately described by a continuous rather than a categorical model of perception.

McClelland and Elman (1986) point out that the categorical output occurs only after some time in which interactive activation has been taking place. Thus, very quick responses might reveal continuous information if the perceiver responds more quickly to the stimulus, as in same–different judgments of the Pisoni and Tash (1974) study. On the other hand, reaction times (RTs) for identification responses are much slower and there should have been sufficient time for interactive interaction. Massaro and Cohen (1983), for example, found that RTs for the identification of bimodal syllables averaged about 1000 ms. Thus, is seems unlikely that the new results supporting continuous rather than categorical perception are due to processing before interactive activation had a chance to take place. In vowel perception, listeners can make within category discriminations even in a task in which the stimuli are not spaced closely together in time. One reasonable explanation is that listeners have better auditory memory for vowels than for consonants (e.g., the second recipe noted earlier). According to interactive interaction, vowels should be perceived as categorically as consonants.

Given the falsification of categorical perception, an important weakness in the TRACE model of speech perception has been documented. It appears that the TRACE model will have to be modified significantly because the evidence points to continuous rather than categorical perception. This modification will not be easy to implement because of the large amount of activation assumed to exist from the phoneme level down to the feature level. It will be interesting to determine if the power of the TRACE model can be maintained without top–down activation of feature evaluation.

Criticisms of Hidden Units

The purpose of this section is to illustrate how (1) connectionist models with hidden units are too powerful, (2) this superpower can camouflage the underlying psychophysics, and (3) and hidden units can camouflage intervening stages of processing. Good models should be falsifiable. However, a single connectionist model can simulate results that imply mutually exclusive psychological processes. Thus, results consistent with a connectionist model should not be taken as evidence for the model. Connectionist models are too powerful. One consequence is that the superpower of hidden units can retard the discovery of the information that a subject uses in a task. Superpowerful models perform adequately even with inappropriate assumptions about the information that is used. Finally, tasks that require relatively successive stages of processing can be simulated to some extent by hidden units. Specifically, a task that requires categorization and response selection stages can be simulated with a layer of hidden units. This simulation is misleading because is seems to imply that the input units have almost immediate contact with the output units. In fact, input processing must usually occur for a significant period before information is made available to the output level. This superpower implies that our standard research strategies will not be sufficient.

The simpliest connectionist models assume only input and output nodes. The stimulus world is mapped into the input nodes, the input nodes are connected to the output nodes, and the output nodes are mapped into the behavioral world. Minsky and Papert (1969) pointed out some of the limitations of these connectionist models, namely, that these models cannot learn to partition events or objects into sets that are not linearly separable from one another. Given objects in a two-dimensional space, for example, the different sets to be categorized differently must be separable by a straight line. Given four objects with the values 00, 01, 10, and 11 along two dimensions, all partitionings are possible except 00 and 11 mapped into one set and 01 and 10 into another. It should be noted that the question whether linear separability is a limitation for psychological models is an empirical one. To overcome the limitations of linear separability, another layer of units must exist between the input and output layers. These hidden units are connected to both input and output nodes.

Superpower of Hidden Units

The present criticism of superpower applies to connectionist models that have hidden units. As anticipated by Minsky and Papert (1969) in their critique of perceptrons (Rosenblatt, 1958), connectionist models with more than two layers of units might be too unconstrained to be informative. Models of this type might be Turing equivalents that are capable of mimicking any computable function. As presently formulated, many of the connectionist models with two-way connections among different levels of units and connectivity among units at a given level appear to be too powerful. They might be capable of predicting not only observed results, but also results that do not occur (Massaro, 1986a). That is, some connectionist models might simulate results that have not been observed in psychological investigations and results generated by incorrect process models of performance.

To test whether a connectionist model with hidden units is too powerful, I examined whether it could simulate data from different, mutually inconsistent processes. Three data sets were generated from three different process models. The data describe the interaction of different sources of information in pattern recognition. The different assumptions made by the different process models are considered to be fundamental to important issues in pattern recognition. If a connectionist model with hidden units can mimic the results of mutually incompatible assumptions about pattern recognition, then the model is too unconstrained to address the issues addressed by the process models.

The three data sets were produced by three different process models of perceptual categorization. The first process model is the fuzzy logical model of perception—a model that has captured a variety of results across a broad range of perceptual and cognitive domains. The other two models are identical to the FLMP in all respects except for the integration rule used to conjoin the multiple sources of information. The second model assumes an additive combination and the third assumes the minimization rule developed in fuzzy logic. These models were used to generate hypothetical data, which were then simulated by connectionist models with hidden units. The goal of these simulations is to test whether connectionist models with hidden units are too powerful.

The three different process models are formulated within a general scheme of pattern recognition, whose major stages are illustrated in Fig. 3. All three models assume these three operations between presentation of a pattern and its categorization. The sources of information are represented by uppercase letters, X_i and Y_j. Each dimension provides a feature value at feature evaluation. The evaluation process transforms each of these sources into psychological values (indicated by lowercase

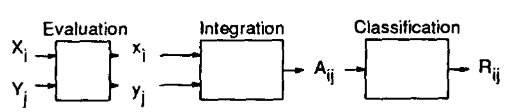

FIG. 3. Schematic representation of three operations involved in perceptual recognition.

letters). Feature evaluation gives the degree to which a given dimension supports each test alternative. The three models differ with respect to the combination rule underlying feature integration: multiplicative, additive, and minimization. The classification operation maps this value into some response, such as a discrete decision or a rating.

To describe the three integration rules, some terminology is necessary. Consider three dimensions of information labeled X, Y, and Z; X_i corresponds to the ith level of the X dimension. Similarly, Y_j and Z_k correspond to the jth and kth levels of the Y and Z dimensions, respectively, A given stimulus composed of a single dimension would be labeled X_i, Y_j, or Z_k, and a given combination of these dimensions would be represented by $X_iY_jZ_k$. The support for some alternative A by dimension X_i is given by the feature value x_i, and analogously for dimensions Y_j and Z_k. Feature integration combines the feature values to determine the total support for a given alternative. If x_i, y_j, and z_k are the values supporting alternative A, then the total support for the alternative A would be given by some combination of x_i, y_j, and z_k. The value A_{ijk} represents the total support given by $x_iy_jz_k$.

For the multiplicative combination rule (which is assumed by the FLMP), total support for alternative A would be given by the product of the feature values:

$$A_{ijk} : x_i \times y_j \times z_k.$$

For the minimization combination rule which is assumed in fuzzy logic (Zadeh, 1965), total support for the alternative A would be given by the minimal value of the feature values:

$$A_{ijk} : \min(x_i, y_j, z_k).$$

For the additive combination rule, total support for the alternative A would be given by the addition of the feature values:

$$A_{ijk} : x_i + y_j + z_k.$$

The third stage of processing assumed by all three models is pattern classification, which gives the relative degree of support for each of the test alternatives. In this case, the probability of an A response given $X_iY_jZ_k$ is

$$P(A|X_iY_jZ_k) = \frac{A_{ijk}}{\Sigma}, \tag{1}$$

where A_{ijk} is given by one of the three combination rules and Σ is equal to the sum of the merit of all relevant alternatives. For each model, the merit of each relevant alternative is derived in the same manner as assumed for alternative A.

Some readers might argue that the three integration rules are not all that different from one another and, therefore, little effort should be invested in distinguishing among the models. More generally, however, two process models of this type can represent highly incompatible assumptions about psychological function. For example, a model can be formalized based on the assumption of selective attention in which only a single source of information is used in the recognition of a test pattern. This model makes predictions similar to the additive model in that no statistical interaction is predicted (Massaro, 1985), and it contrasts sharply with the FLMP which assumes that all three sources of information influence the perceptual recognition of a given test pattern. Even readers who interpret the three integration rules as similar to one another will have to admit that the selective attention model is fundamentally different from an integration model.

The data generated for the simulations were generated from an expanded three-factor design. The expanded design tests each of the three factors presented alone, as well as the factorial combination of all three factors. The design provides a more powerful database to assess models of human performance than standard factorial designs (Massaro, 1987b). There were seven levels of each of the three variables. Each process model predicts that the probability of a particular identification is some combination of unique parameter values associated with each of the levels of the three independent variables. All three process models assume that the parameter values are between zero and one. Predictions can therefore be made for each of the process models from the same set of parameter values for each of the three variables.

The parameter values used in generating the predictions are given in Table 1. The predictions of the three models, given these parameter values, are given by the points in Figs. 4, 5, and 6, respectively. The left panel of each figure gives the observations for the interaction of variable Y and variable Z, when variable X is equal to 0.5. The right panel gives the three single-variable conditions. As can be seen in the figures each of the models makes a different prediction. Especially noticeable is the difference between the adding combination rule and the other two models. The differences between the multiplicative integration rule and the minimization rule are also apparent in the form of the interaction between variable Y and variable Z. Before testing the connectionist model against these three results, it is informative to evaluate how much the three different process models differ from one another. It is logically possible that one process model can mimic the results of another, simply with a change in parameter values. To explore this issue, the three process models were fit to the three sets of results generated by these same process models. In all cases, 21 parameter values (3 variable × 7 levels for each variable) were estimated to minimize the differences between the "observed" and predicted data. The criterion of best fit was based on the root mean square deviation (RMSD) or the squared root of the average squared deviation between predicted and observed points. As can be seen in Table 2, the models can describe data consistent with their assumption, but cannot describe data generated by the other process models. The RMSD values for "incorrect" models are sufficiently large enough to warrant the belief that these process modles could be distinguished from one another in practice.

TABLE 1
ORIGINAL PARAMETER VALUES USED TO GENERATE THE PREDICTIONS OF THE THREE PROCESS MODELS IN THE EXPANDED FACTORIAL DESIGN

	Level						
Dimension	1	2	3	4	5	6	7
X	.010	.100	.300	.500	.700	.900	.990
Y	.001	.200	.400	.600	.800	.970	.999
Z	.005	.170	.250	.750	.860	.940	.980

A connectionist model with six hidden units was created using the framework given by Rumelhart, Hinton, and Williams (1986). A unique input node was assumed to be activated by each unique level of each of the three independent variables, giving a total of 3 × 7 = 21 input nodes. Table 3 gives a schematic representation of this input architecture. A given stimulus was assumed to activate the nodes corresponding to the levels of the variables contained in the stimulus. Thus, a single node was activated for the single variable presentations and three nodes were activated for the three-factor presentations. Activation of a given node was assumed to produce an output of one for that unit. Otherwise the output was zero. There were also two output units corresponding to the two hypothetical response alternatives in the task. With six hidden units, there are 21 × 6 = 126 weights that must be determined to give the connection strengths between the input and hidden units. In addition, there are 6 × 2 = 12 weights connecting the hidden to the output units. Eight weights are necessary for the threshold units connected to the hidden and output units. This given a total of 146 weights to predict the 364 experimental conditions.

This connectionist model with six hidden units was fit to the predictions of each of the three process models. For each set of

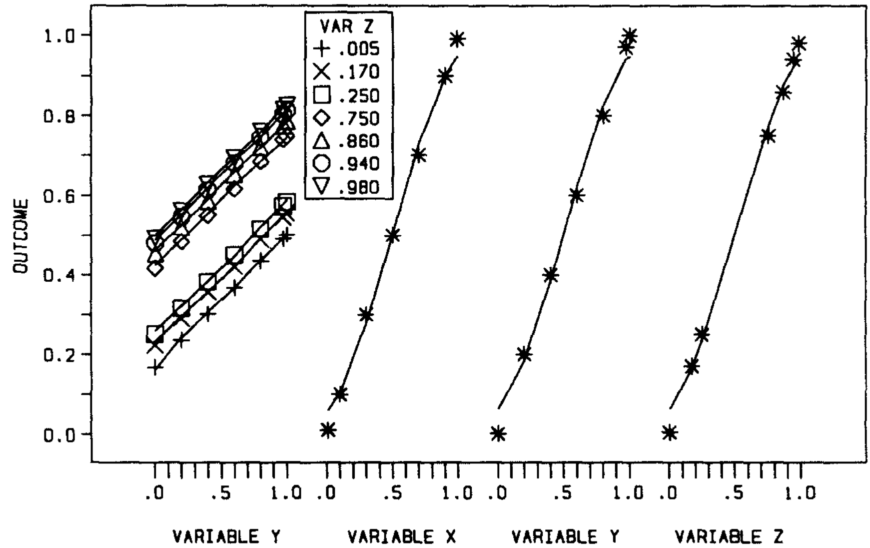

FIG. 4. Data (points) generated by the additive process model and predictions (lines) of a connectionist model with six hidden units and the architecture given in Table 3.

data, weights were determined to give the best possible predictions of the data. The weights were determined by either backpropagation (Rumelhart et al., 1986) or STEPIT (Chandler, 1969; Massaro, 1987b). We have found that these two parameter estimation techniques give similar fits of data, even though they are substantially different from one another. STEPIT and other minimization techniques based on variants of Newton's method make adjustments in the parameter values based on a global index of goodness of fit. Backpropagation, on the other hand, makes adjustments in parameter values based on a local index of goodness of fit. As can be seen in Table 4 and Figs. 4 through 6, the connectionist model can give a good description of

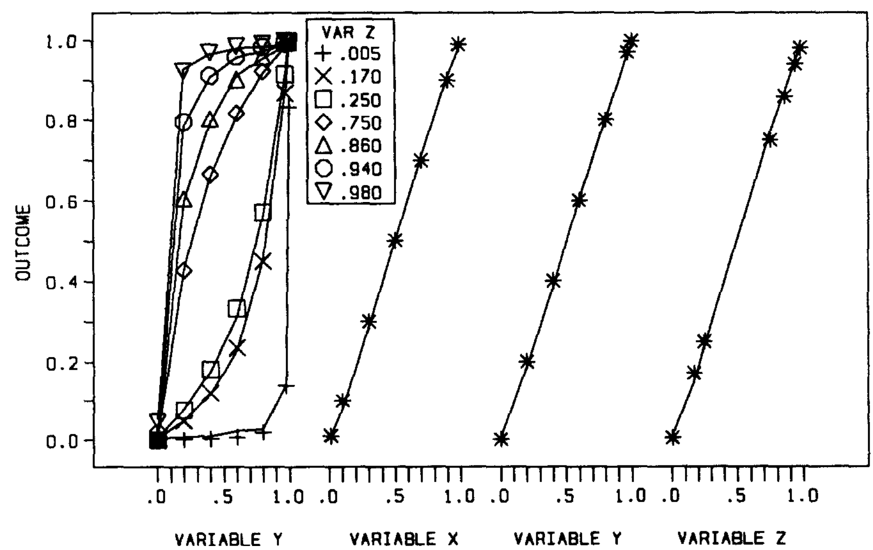

FIG. 5. Data (points) generated by the multiplicative process model and predictions (lines) of a connectionist model with six hidden units and the architecture given in Table 3.

all three sets of data. A good description was defined as an RMSD of 1% which was also the terminating criterion in the actual fit of the connectionist models. Given these results, the model appears to be superpowerful or capable of predicting three different types of results that were predicted by three different process models.

To show that the fit of the model is critically dependent on the specific weights determined in training, the connectionist model with weights determined from the fit of the additive data was tested against the data generated by the other two process models. Carrying out this analysis for the three sets of weights and the three data sets gives nine independent tests. Table 4 shows that the weights are critical to achieve a good fit. The weights determined to maximize the description of one set of data give a poor description of another set of data.

An immediate reaction to my demonstrations might be that I am doing curve fitting whereas backpropagation is meant to do learning. The conclusions implicated by my analyses, therefore, cannot be extrapolated to studies of learning. My reply is that modeling learning and curve fitting are very similar in practice and a sharp distinction between the two is not possible. The teaching input or target activation in backpropagation is analogous to the observed data in curve fitting. In both cases, the goal is to minimize the deviations between two sets of values. The fact that backpropagation has been used to learn on some training data and then tested on new test data is not a distinguishing characteristic; curve fitting can be carried out in a similar manner. Rather than be bogged down with this issue, however, allow me to frame the exercise as follows. Are there unique sets of weights that allow uniquely different data sets to be predicted by the same model? If there are, I take this as an instance of superpower because how the

TABLE 2

RMSD VALUES FOR THE FIT OF THE THREE PROCESS MODELS TO THE DATA GENERATED BY THE THREE PROCESS MODELS IN AN EXPANDED THREE-FACTOR DESIGN

	Data		
Model	Add	Mult	Min
Add	.0000	.2579	.2473
Mult	.0578	.0000	.0651
Min	.0713	.1220	.0000

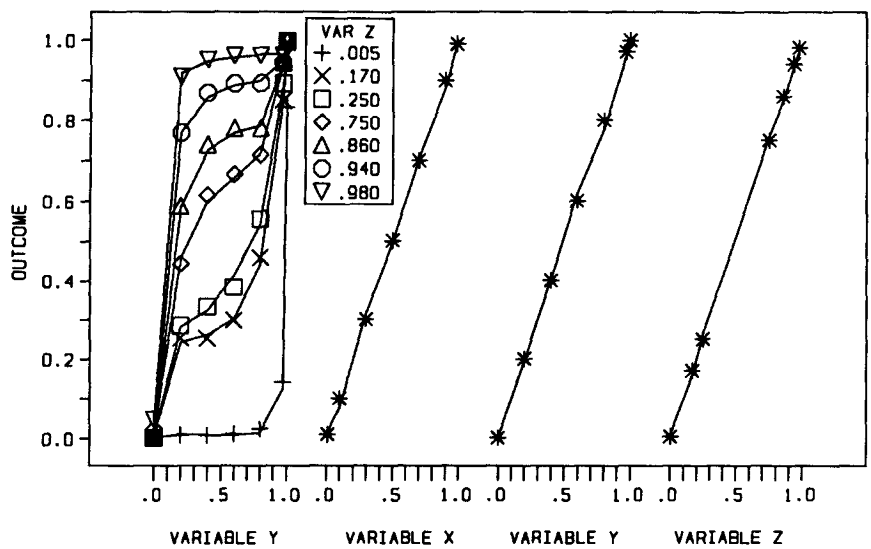

FIG. 6. Data (points) generated by the minimization process model and predictions (lines) of a connectionist model with six hidden units and the architecture given in Table 3.

weights are determined is secondary. That is, without any strict a priori assumptions about learning in the situation of interest, the ingenious investigator will come up with a plausible learning history to justify the weights that are desired.

The procedure used in the present investigation differs from that used in previous studies using connectionist models. In those studies, the parameter values are either crafted by hand (McClelland & Rumelhart, 1981; McClelland & Elman, 1986) or derived from an initial session of learning to identify a set of patterns (McClelland & Rumelhart, 1985). An obvious criticism of the present work is that a connectionist model is revealed to be superpowerful only because the weights are being determined directly from the results being predicted. That is, we are estimating the weights from the same data that is being described, whereas one should determine the weights without reference to the results being predicted. Given this procedure, one might argue that I cannot generalize from the negative conclusion reached in the present paper. However, my reply would be that there have been very few constraints on the handcrafting of the parameters and on the learning conditions, and therefore it is not obvious that the present strategy is all that different from what has gone on previously. The main difference is the criterion used for goodness-of-fit. The RMSD criterion has been used to seek a close quantitative fit, whereas previous tests of connectionist models have been concerned primarily with generating qualitative results paralleling those of interest.

The current endeavor might be modified slightly to resemble the earlier approaches by fitting new sets of data with the same weights. In this case, the question of interest is whether training on additive data will give accurate predictions to new sets of additive data, and so on. We might expect that the learning experience only has to be taylored to the test for the model to produce a good match to the results. To carry out this analysis, we took the weights derived in the fit of a particular data set, tested each model against the data set, and applied the connectionist model with these weights to the two-factor predictions of the three process models. We did not include the pairwise combinations of the three independent variables in the original expanded factorial design. These two-factor combinations give 3 × 49 = 147 additional data points to be predicted by using the weights that were determined in fitting the original expanded factorial design.

The connectionist model was tested against these new data using the weights derived in the original fit. With three sets of weights and three sets of two-factor data,

TABLE 3
INPUT NODE ACTIVATIONS AND ASSOCIATED PARAMETER VALUES FOR EACH OF THE THREE VARIABLES, GIVEN THE SEVEN-NODE ARCHITECTURE

Input node activations	Value X	Value Y	Value Z
■□□□□□□	.010	.001	.005
□■□□□□□	.100	.200	.170
□□■□□□□	.300	.400	.250
□□□■□□□	.500	.600	.750
□□□□■□□	.700	.800	.860
□□□□□■□	.900	.970	.940
□□□□□□■	.990	.999	.980

Note. Each level of each variable is assumed to activate one of the seven unique nodes associated with that variable.

TABLE 4
RMSD VALUES FOR THE FIT OF A CONNECTIONIST MODEL WITH A SEVEN-NODE ARCHITECTURE AND SIX HIDDEN UNITS

	Data tested		
Data trained	Add	Mult	Min
Add	.0100	.2617	.2536
Mult	.2592	.0100	.1408
Min	.2539	.1412	.0100

Note. The data were generated by three different process models in an expanded factorial design. The network was trained on each of the three sets of data and tested on each of the three sets of data.

there are nine conditions. The superpower is maintained even though new weights were not estimated. Table 5 gives the RMSD values for the different conditions. As can be seen in the Table 5, the connectionist model with weights derived from the fit of additive data gave a good description of new two-factor data that were generated from the additive process model. Similar results were found for the other two data sets. The connectionist model does a good job describing a new set of data if the new set of data is consistent with the data that were originally used to train the connectionist system.

In summary, a specific connectionist model with hidden units was shown to be superpowerful in a specific experimental situation because it could predict three different types of results generated by three different process models. Good scientific inquiry dictates that we actively attempt to eliminate alternative models. Given the power of connectionism, however, the observed data often will not be sufficient to decide among alternative models. Luckily, there is another strategy that might be employed beneficially. In this approach, only models with discriminating taste would be permitted to survive. Discriminating taste means that a model *only* predicts actual results, not the universe of possible results. The investigator therefore addresses not only observed results but also a range of hypothetical results that *do not* necessarily occur. Collyer (1985) has made a similar point that although a more complex model might be more accurate than a less complex model, the more complex model should not necessarily be preferred.

TABLE 5
RMSD VALUES FOR THE FIT OF A CONNECTIONIST MODEL WITH A SEVEN-NODE ARCHITECTURE AND SIX HIDDEN UNITS

Add	Mult	Min
.0317	.0150	.0358

Note. The data were the two-variable combinations of an expanded three-factor design. The weights that were used were derived from the fit of the model to the data generated by three different process models in an expanded factorial design.

The three different process models were not superpowerful; each process model could not predict results generated by its competitors. Not all connectionist models are superpowerful; in fact, a two-layer connectionist model can be formulated to be mathematically equivalent to the FLMP in some situations (Massaro & Cohen, 1987). This connectionist model has discriminating taste; it predicts the results of actual results, but not the results generated by other process models.

Compromising Psychophysics

Central to connectionist models is a featural representation of the input, such as the Wickelphone representation in the verb-learning model (Rumelhart & McClelland, 1986). A good description by the model is taken as support for both the model and the featural representation. However, the superpower of connectionist models with hidden units can mislead us about the validity of a given featural representation and hence about psychophysical relationships. Psychophysics refers to the relationship between a stimulus domain and a sensory/perceptual domain. As an example, the stimulus domain might be letters of the alphabet printed in a Helvetica type font and the psychological domain certain visual features that are functional in letter recognition. For purposes of the present demonstration, we define *appropriate* and *inappropriate* featural representations. In the appropriate representation, the features assumed by the connectionist model are equivalent to those used by the subject in the task. In the inappropriate representation, the features assumed by the connectionist model differ significantly from those used by the subject in the task.

A falsifiable model should not behave equally well with appropriate and inappropriate featural representations. To show the limitations of superpowerful models with hidden units, I generated one featural representation that was appropriate and another featural representation that was inappropriate. In the appropriate representation, the active nodes were systematically related to the functional stimulus. Each independent variable was designed to have a unidimensional relationship with the categorization of the stimulus. We assume that there are nine features and adjacent stimuli would always have more features in common than nonadjacent stimuli. Thus, adjacent stimuli along the continuum would be most similar, and this similarity relationship was preserved in the nine-node representation, as illustrated in Table 6. Adjacent stimuli shared two active nodes in common; stimuli separated by one stimulus shared one active node in common; and stimuli separated by two or more nodes had no active nodes in common.

In the inappropriate representation, I designed an irregular relationship between the featural representation and the categorization of the stimulus, as illustrated in Table 7. As can be seen, adjacent stimuli sometimes have two, one, or no active nodes in common, and this is also true for nonadjacent stimuli. In this case, the featural representation of the connectionist model actually violates the information that is being used by the subject.

Fifteen hidden units were chosen for the three-feature representation and six units for the nine-node representation to equate the number of weights (or free parameters) for the two models. The nine-node model with six hidden units has 182 weights, as does the three-node model with 15 hidden units.

The learning history for these tests of the connectionist models might be expected to provide valuable information about the adequacy of the models. However, variables such as the number of trials to achieve a given goodness-of-fit criterion are highly dependent on the learning parameters. In addition, both backpropagation and STEPIT sometimes fail by settling into nonoptimal local minima. Given the computing power and time needed to test these models, it is not feasible to provide a detailed investigation of all of the contributing factors. Once again, we have to be content with the final fit that is achieved rather than how we achieved it. If an investigator is interested in models of learning, my belief is that they must be analyzed and tested against trial-by-trial data.

The appropriate and inappropriate connectionist architectures were fit to the three sets of data generated by the three process models by estimating weights to minimize the RMSD. The predictions of

TABLE 6

INPUT NODE ACTIVATIONS AND ASSOCIATED PARAMETER VALUES FOR EACH OF THE THREE VARIABLES, GIVEN AN APPROPRIATE NINE-NODE ARCHITECTURE

Input node activations	Value *A*	Value *B*	Value *C*
■■■□□□□□□	.010	.001	.005
□■■■□□□□□	.100	.200	.170
□□■■■□□□□	.300	.400	.250
□□□■■■□□□	.500	.600	.750
□□□□■■■□□	.700	.800	.860
□□□□□■■■□	.900	.970	.940
□□□□□□■■■	.990	.999	.980

Note. Each level of each variable is assumed to activate three of the nine unique nodes associated with that variable.

TABLE 7

INPUT NODE ACTIVATIONS AND ASSOCIATED PARAMETER VALUES FOR EACH OF THE THREE VARIABLES, GIVEN AN INAPPROPRIATE THREE-NODE ARCHITECTURE

Input node activations	Value*X*	Value *Y*	Value *Z*
□□■	.010	.001	.005
□■□	.100	.200	.170
□■■	.300	.400	.250
■□□	.500	.600	.750
■□■	.700	.800	.860
■■□	.900	.970	.940
■■■	.990	.999	.980

Note. Each level of each variable is assumed to activate a unique configuration of the three nodes associated with that variable.

the FLMP, given the parameter values in Table 1, and representative simulations for the nine-node and three-node architectures are shown in Figs. 7 and 8, respectively. As can be seen in the figures, a model with hidden units with the inappropriate or abnormal three-node representation does about as well as a model with the appropriate nine-node representation. Given that adequate performance of a model is taken to provide evidence for both the model and the featural representation, we see that the superpower of models with hidden units can mislead us not only about the models themselves but also about the relevant psychophysics. As long as connectionist models are unconstrained and Turing equivalents in principle, research in the field will find it difficult to address fundamental psychophysical questions. What are the features of letters and words actually used in reading? It does not really seem to matter for connectionist models. Nonveridicality on the input side can be compensated for on the processing side. Adding more hidden units or allowing more connections among units could overcome deficiencies in representation on the input side.

All things being equal, a more valid representation should give a better description of the results. To illustrate this point, the number of hidden units was reduced in the nine-node and three-node representations. Two hidden units were employed in the regular nine-node architecture and five hidden units were assumed for the irregular architecture. These models require roughly the same number of weights: 66 for the regular and 62 for the irregular architectures. The regular architecture gave a better description of the results than did the irregular architecture, with RMSDs of .01 and .04, respectively. This demonstration offers a procedure that might be valuable in testing different models and representations. In general, investigations would gain in credibility if the preferred model gave a better fit than nonpreferred models with the same preferred representation, and if the preferred model gave significantly poorer fits with nonpreferred representations.

It is critical that connectionist models work out the psychophysics hand-in-hand with the information processing that occurs. I have argued elsewhere (Massaro, 1987b) that both the frameworks of psychophysics (specifying the environmental

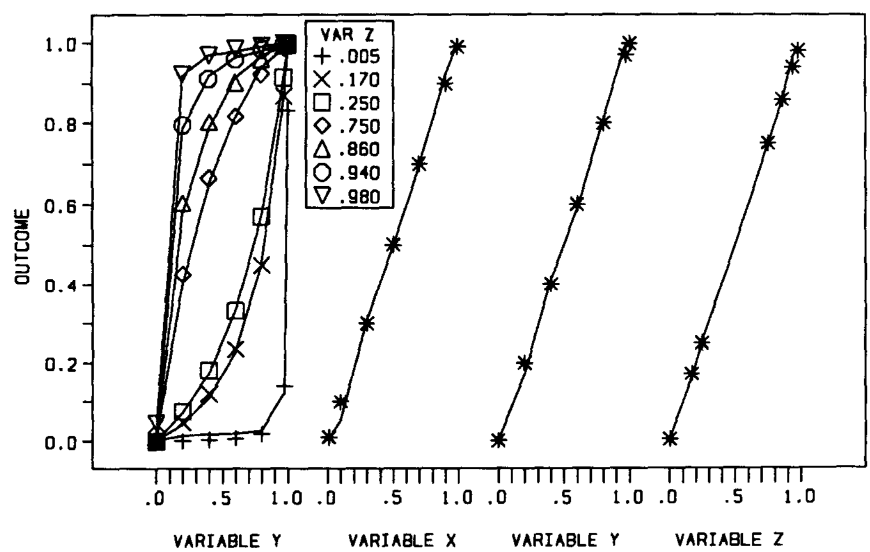

FIG. 7. Data (points) generated by the multiplicative process model and predictions (lines) of a connectionist model with six hidden units and the appropriate architecture given in Table 6.

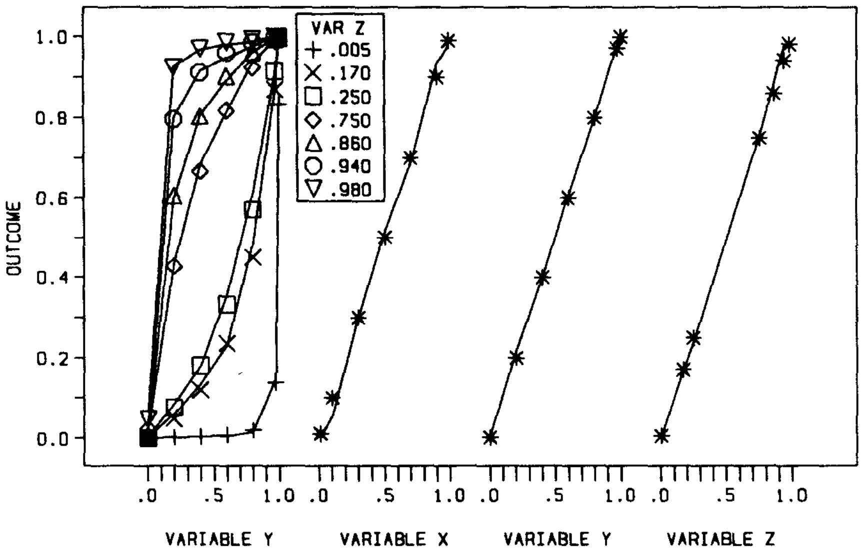

FIG. 8. Data (points) generated by the multiplicative process model and predictions (lines) of a connectionist model with 15 hidden units and the inappropriate architecture given in Table 7.

characteristics that are utilized by subjects) and information processing (specifying the processing that these characteristics undergo) are necessary for progress in the field. Apparently, connectionist models do not work out the psychophysics in any great detail. In fact, the claims that they make are often compromised because of the acknowledgment of the arbitrary nature of the psychophysics that they assume. As an example, in the TRACE model of McClelland and Elman (1986), they acknowledge that although the features, phonemes, and words levels were central to the predictions, these may not be the actual units that are functional. They go on to state that their model is equally valid even if different units are, in fact, functional.

Compromising Stages of Processing

The superpower of connectionist models with hidden units can also camouflage the contribution of different stages of processing. As an example, consider two tasks involving the letters of the alphabet. In one task, subjects are asked to identify the letters and, in the other, they are asked to identify and pronounce the letters. In the first task, both the input and output representations could be in terms of visual features. In the second task, the input representations could be visual features but the output representations should be articulatory ones. An information-processing analysis of the second task would impute an additional stage of categorization before pronunciation is selected. Connectionist models, on the other hand, do not have intervening stages of processing, but simply a mapping between input and output. Within the connectionist framework, the two tasks differ in terms of the goodness of this input–output mapping. A "good" mapping refers to linearly separable data in which a linear discriminant function can achieve a correct mapping of inputs to output categories. The output category can be determined by a sum of weighted values on each of the elements of the input. The "bad" mapping does not have a solution in terms of a linear discriminant function.

With respect to the two tasks involving letters, the identification task corresponds to a good mapping and the pronunciation task corresponds to a bad mapping. The mapping between input and output would be linearly separable for the identification task, but not for the pronunciation test. If the input consists of visual features of the letters and the output consists of articulatory features of the pronunciation, it is unlikely that the mapping between input and output would be linearly separable. For example, the letters *c* and *o* would have more similar input representations than the letters *c* and *z*. However, the output representations would be similar for *c* and *z* than for the letters *c* and *o*. Some intermediate categorization is necessarily involved to map visual properties of the letters to articulatory properties of the pronunciation. In experimental situations somewhat analogous to this example, Miller (1982) found that some type of categorization occurs before input is mapped into output.

The superpower of connectionist models with hidden units can camouflage the important differences between the identification and pronunciation tasks. With enough hidden units, the two tasks can be described equally well with the same architecture and the same input–output representations. To illustrate this thesis, I generated a set of good data and a set of bad data in terms of input–output mappings. Table 8 gives the input and output structures for the two mappings. Three input nodes are one or zero and one of the four output nodes is one. These data were fit with a two-layer connectionist model with no hidden units, as well as with models with hidden units. There were either 2, 5, or 10 hidden units at a single hidden layer or there were two layers of 5 hidden units each.

Table 9 shows that both the two-layer model and the models with hidden units fit the good data perfectly. For the bad data, only the models with 10 hidden units (whether in two hidden layers or one) could describe the mapping accurately. Solving the mapping for the bad data depends only on the number of hidden units (free parameters) and does not seem to illuminate the relationship between input and output. The hidden units serve only to bypass the assumption of an intervening stage of processing in which the input is categorized before an output is selected.

Categorization is a natural intervening process that is assumed in information processing but not in current connectionist models. For example, NETtalk (Sejnowski & Rosenberg, 1986) is a connectionist model aimed at translating written language into spoken language. It is only natural to expect that some intermediate categorization of the letters or words would necessarily occur between some input representation of the written text and some output representation of the pronunciation. However, this type of categorization is not easily implemented in the connectionist model, and falls on the shoulders of hidden units. Thus, NETtalk's architects probably settled on letters rather than visual features as input because the number of hidden units to solve the mapping of visual fea-

TABLE 8
INPUT NODE ACTIVATIONS AND ASSOCIATED RESPONSES GIVEN THE GOOD AND BAD CONFIGURATIONS

Input node activations	Good	Bad
□□■	0001	0100
□■□	0010	0010
□■■	0001	0001
■□□	1000	0001
■□■	0100	0010
■■□	0010	0100
■■■	0100	1000

TABLE 9
RMSD VALUES FOR THE FIT OF THE THREE PROCESS MODELS TO GOOD DATA THAT ARE LINEARLY SEPARABLE AND BAD DATA THAT ARE NOT LINEARLY SEPARABLE

	Data	
Model	Good	Bad
Direct	0	.327
2 hidden	0	.269
5 hidden	0	.134
5 × 5 hidden	0	0
10 hidden	0	0

tures to pronunciation would have been prohibitive. On the other hand, visual features could have been used if some process of letter categorization intervened between the visual features and pronunciation. My belief is that connectionism will have to become more stage-like to solve input–output mappings in an informative manner.

Relationship of Connectionist to Process Models

A prototypical connectionist framework shares several fundamental properties with an information-processing model of pattern recognition, as instantiated in the FLMP (Massaro, 1986a, 1986b; Oden & Rueckl, 1986). First, both frameworks assume multivalued (continuous) rather than binary (discrete) representations; the fuzzy truth values of the FLMP are analogous to the continuous levels of activation and inhibition of connectionist models. Second, both frameworks acknowledge the existence of multiple simultaneous constraints on human performance. Both frameworks provide an account of the evaluation and integration of multiple sources of information in pattern recognition. Third, there is the parallel assessment of multiple candidates or hypotheses at multiple levels in both models. Fourth, both frameworks provide a common metric for relating qualitatively different sources of information. In the FLMP, each source of information is represented by fuzzy truth values representing the degree to which alternative hypotheses are supported. Activation levels of sets of units play the analogous role in connectionist models. Fifth, the automatic categorization of a novel instance can be accomplished in both frameworks. Finally, both frameworks conceptualize pattern recognition as finding the best fit between the relevant constraints and the pattern that is perceived.

The close fit between the present framework and connectionism dictates an exploration of their similarities and differences. Although the two frameworks appear to agree on important theoretical criteria, the specific models to date differ in terms of the amount of connectivity in the system. The FLMP assumes no top–down influences of a higher-level unit on activation of a lower-level unit and no inhibition among units at a given level. Connectionist models, such as the interactive activation models of written word recognition and speech perception, usually make both of these assumptions. The analyses presented earlier in this paper question these assumptions in that they are either wrong or unnecessary. Analogous to models with hidden units, connectionist models with two-way connections among different levels of units and connectivity among units at a given level are too powerful. They are capable of predicting not only observed results but also results that do not occur (Massaro, 1986a).

In a recent paper, we proved that a particular two-layer connectionist model (CMP) is mathematically equivalent to the FLMP (Massaro & Cohen, 1987), in a situation with two response alternatives. The two-layer connectionist model is limited to input and output units, with connections only from input to output. The FLMP specifies evaluation of a source of information in terms of a truth value representing the degree to which an alternative is supported. The integration of different sources of information is specified by the multiplication of the truth values. The CMP implements evaluation and integration by activations and inhibitions between an input layer of units and an output layer of units. Evaluation of a source of information corresponds to the activation along a single connection between an input unit and an output unit. Integration in the CMP corresponds to the sum of all the activations entering a given output unit, and transformed by the sigmoid squashing function. Both models assume that a decision is made on the basis of the relative goodness of match, as dictated by Luce's (1959) choice rule. The mathematical correspondence between the FLMP and CMP reveals that the two models, couched in different theoretical frameworks, can make identical predictions in practice. Furthermore, these predictions are consistent with quantitative results in several domains of pattern recognition (Massaro, 1984a, 1984b, 1987b; Oden, 1978, 1981). The similarity of a connectionist model without interactive activation or hidden units to the FLMP shows that process and connectionist models are not necessarily incompatible.

Retrospective

My purpose in writing this paper is not to cast a pejorative view of connectionism, but to offer strategies for inquiry that would enhance its contribution to experimental and theoretical psychology. While completing this paper, I intercepted a very recent critique of cognitive neuropsychology written by Ellis (1987). This new domain of inquiry attempts to relate specific cases of brain trauma to disruption or elimination of specific information-processing modules. His observations of the current state of the art in that field capture several of the themes of the present paper. He points out some trends in that field which could bring about its the second demise; analogously, I have criticized current strategies in connectionism that could lead to its second downfall. Ellis also argues against the apparent move toward a separate discipline apart from cognitive experimental psychology. I believe that it would be equally disastrous for connectionism to become a discipline of its own.

Most research within connectionism has not yet addressed human performance in the manner described by my suggestions for good, scientific inquiry (Massaro, 1986a). Tests of alternative models are not usually presented. The prototypical study involves presentation of a model that works with little discussion of what does not work and whether a simpler model would work as well. As a promising exception, Golden (1986) contrasted two connectionist models of visual word perception. Golden's (1986) model and Brown's (1987) model are examples that the interactive activation model was much more powerful than necessary. Furthermore, history repeats itself in that we do not know if the latter two models are too powerful also. In addition, seldom is a fine-grained analysis of the results performed. Usually, the performance of the model is compared to the performance of human subjects at only a qualitative level.

In all cases, faced with these criticisms, one can see how they might be addressed in practice; however, it remains to be seen if investigators in the connectionist framework will modify their research strategy accordingly. Given the influence of connectionism on current endeavors in experimental psychology, we can expect to see much more rigorous study as experimental psychologists with years of good, scientific training do what they know best. When this occurs, my prediction is that connectionist models will become more like information-processing models. Superpowerful models will be discarded. The remaining class of models will not be able to predict at a fine-grain level the mapping of perception into action without fairly veridical psychophysics and something like stages of processing.

In summary, the information-processing paradigm will probably be shaped by some aspects of both modularity and connectionism. More importantly, we can now see that connectionism does not represent a Kuhnian paradigm shift (Kuhn, 1962; Schneider, 1987) because if follows naturally from the paradigm of information processing. Progress in psychological science, as in other sciences, is continuous and not catastrophic. There is no shortage of models of human performance, and neural constraints are not currently available to reduce the number of alternatives. The more things change, the more they remain the same. Much work remains to be done.

REFERENCES

BROWN, G. D. A. (1987). Resolving inconsistency: A computational model of word naming. *Journal of Memory and Language*, **26**, 1–23.

CHANDLER, J. P. (1969). Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, **14**, 81–82.

COLLYER, C. E. (1985). Comparing strong and weak models by fitting them to computer-generated data. *Perception and Psychophysics*, **38**, 476–481.

ELLIS, A. W. (1987). Intimations of modularity, or, the modelarity of mind: Doing cognitive neuropsychology without syndromes. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language* (pp. 397–408. Hillsdale, NJ: Erlbaum.

FODOR, J. A. (1983). *Modularity of mind*. Cambridge, MA: Bradford Books.

GOLDEN, R. M. (1986). A developmental neural model of visual word perception. *Cognitive Science*, **10**, 241–276.

HARY, J. M., & MASSARO, D. W. (1982). Categorical results do not imply categorical perception. *Perception and Psychophysics*, **32**, 409–418.

HULL, C. L. (1943). *Principles of behavior*. New York: Appleton–Century–Crofts.

KUHN, T. S. (1962). *The structure of scientific revolutions*. Chicago: Univ. of Chicago.

LUCE, R. D. (1959). *Individual choice behavior*. New York: Wiley.

MASSARO, D. W. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, **5**, 595–609.

MASSARO, D. W. (1984a). Building and testing models of reading processes. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 111–146). New York: Longman.

MASSARO, D. W. (1984b). Time's role for information, processing, and normalization. *Annals of the New York Academy of Sciences, Timing and Time Perception*, **423**, 372–384.

MASSARO, D. W. (1985). Attention and perception: An information-integration perspective. *Acta Psychologica*, **60**, 211–243.

MASSARO, D. W. (1986a November). *Connectionist models of mind*. Paper given at the Twenty-Seventh Annual Meeting of the Psychonomic Society, New Orleans.

MASSARO, D. W. (1986b). The computer as a metaphor for psychological inquiry: Considerations and recommendations. *Behavior Research Methods, Instruments, & Computers*, **18**, 73–92.

MASSARO, D. W. (1987a). Categorical partition: A fuzzy-logical model of speech perception. In S. Harnad (Ed.), *Categorical perception* (pp. 254–283). New York: Cambridge Univ. Press.

MASSARO, D. W. (1987b). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

MASSARO, D. W., & COHEN, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, **9**, 753–771.

MASSARO, D. W., & COHEN, M. M. (1987). *Process and connectionist models of pattern recognition*. Proceedings of the Ninth Annual Conference of the Cognitive Science Society (pp. 258–264). Hillsdale, NJ: Erlbaum.

MASSARO, D. W., & ODEN, G. C. (1980). Speech perception: A framework for research and theory. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 3, pp. 129–165). New York: Academic Press.

MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1–86.

MCCLELLAND, J. L., & RUMELHART, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, **88**, 375–407.

MCCLELLAND, J. L., & RUMELHART, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, **114**, 159–188.

MCCLELLAND, J. L., & RUMELHART, D. E. (1986). *Parallel distributed processing, Vol. 2: Psychological and biological models*. Cambridge, MA: MIT Press.

MCGURK, H., & MACDONALD (1976). Hearing lips and seeing voices. *Nature(London)*, **264**, 746–748.

MILLER, J. (1982). Discrete verus continuous stage models of human information processing: In search of partial output. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 273–296.

MINSKY, M., & PAPERT, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

NEISSER, U. (1967). *Cognitive psychology*. New York: Appleton–Century–Crofts.

ODEN, G. C. (1978). Semantic constraints and judged preference for interpretations of ambiguous sentences. *Memory & Cognition*, **6**, 26–37.

ODEN, G. C. (1981). A fuzzy propositional model of concept structure and use: A case study in object identification. In G. W. Lasker (Ed.), *Applied* *systems and cybernetics* (Vol. VI, pp. 2890–2897). Elmsford, NY: Pergamon.

ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172–191.

ODEN, G. C., & RUECKL, J. G. (1986, November). *Taking language by the hand: Reading handwritten words*. Paper given at the Twenty-Seventh Annual Meeting of the Psychonomic Society, New Orleans.

PALMER, S. E., & KIMCHI, R. (1986). The information processing approach to cognition. In F. J. Knapp & L. C. Robertson (Eds.), *Approaches to cognition: Contrasts and controversies* (pp. 37–77). Hillsdale, NJ: Erlbaum.

REPP. B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 10, pp. 243–335). New York: Academic Press.

ROSENBLATT, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**, 386–407.

RUMELHART, D. E., HINTON, G. E., & WILLIAMS, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Vol. 1. Foundations*. Cambridge, MA: MIT Press.

RUMELHART, D. E., & MCCLELLAND, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, **89**, 60–94.

RUMELHART, D. E., & MCCLELLAND, J. L. (Eds.) (1986). *Parallel distributed processing: Vol. 1. Foundations*. Cambridge, MA: MIT Press.

SCHNEIDER, W. (1987). Connectionism: Is it a paradigm shift for psychology? *Behavior Research Methods, Instruments, & Computers*, **19**, 73–83.

SEJNOWSKI, T. J., & ROSENBERG, C. R. (1986). *NETtalk: A parallel network that learns to read aloud*. The Johns Hopkins University Electrical Engineering and Computer Science Technical Report, JHU/EECS-86/01.

WATSON, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, **20**, 158–177.

ZADEH, L. A. (1965). Fuzzy sets. *Information and Control*, **8**, 338–353.

(Received June 12, 1987)
(Revision received November 2, 1987)