Psychol Res (1986) 48:123–132

Psychological Research
© Springer-Verlag 1986

Addressing issues in letter recognition

Dominic W. Massaro and Joseph M. Hary

Program in Experimental Psychology, University of California, Santa Cruz, CA 95064, USA

Summary. The present research strategy utilizes factorial designs, functional measurement, testing of mathematical models and strong inference in the study of letter perception. To test the viability of this framework, subjects judged a number of ambiguous letters, varying between Q and G, in both a rating and discrete choice task. The letters were created by varying features of openness in the oval and the obliqueness of the straight line. Experimental and theoretical tests on the results indicate that multiple sources of featural information simultaneously contribute to the perception of letters. The features provide continuous rather than discrete information to an integration process and the evaluation of the information provided by one feature is independent of the nature of the other features. The integration process results in the least ambiguous letter feature contributing the most to the perceptual judgment. A fuzzy logical model developed in other domains, such as speech perception, provides a good description of exactly these phenomena.

Introduction

Although scientists have been engaged in reading-related research for a century or so, we have yet to formulate fundamental laws about reading. We have learned, though, that word recognition is a fundamental component of reading and learning to read. Studies of reading ability, eye movements, naming, and lexical decision tasks establish the importance of word recognition in reading (Just & Carpenter, 1980; Carrithers & Bever, 1984; Gough, 1984; Perfetti, Goldman, & Hogaboam, 1979). Not only is word recognition fundamental to reading, but letter recognition is the basis of word recognition. Higher order perceptual cues, such as context and overall word shape, cannot account for word recognition; there is good evidence that word recognition is mediated by the resolution of the letters composing a word (Gough, 1984; Massaro, 1984; Paap, Newsome, & Noel, 1984).

Our perspective, then is that understanding the processes involved in letter recognition is crucial to our understanding of word recognition and ultimately reading. Traditionally, researchers have taken two approaches to the study of letter recognition. One method is to examine the pattern of errors that subjects make when they identify letters (Bouma, 1971; Cattell, 1886; Loomis, 1982). To induce a reader to make errors, letter stimuli are degraded by presenting them for a short duration or at a great distance. The responses of the subjects are entered into a confusion matrix which indicates the identification given to each letter stimulus. For example, subjects might be given the set of 26 lowercase letters in English and they respond with one of the 26 alternatives on each trial. These results are then used to distinguish among various descriptions of the properties of the letters. The goal has been to find the smallest set of properties that best describes the responses. The most popular method of data analysis in this domain has been multidimensional scaling (Gilmore, Hersh, Caramazza, & Friffin, 1979; Townsend, 1971). Although the relative similarity of the letters might be described, one limitation of this general method is that the nature of the psychological processes can not be determined.

The second method of study, the approach that we take in this paper, involves the systematic manipulation of the properties of letters. Subjects identify letters modified in systematic ways, and their responses are used to test quantitative models of the identification process (Oden, 1979; Naus & Shillman, 1976). An important distinction must be made between the visual characteristics of the letters that are manipulated in the experiment and the visual features that the readers actually utilize in the identification of the letters (Massaro & Schmuller, 1975, p. 209). Letters can be described by an almost endless number of characteristics or properties (Palmer, 1978) and only a small set of these will be psychologically real. Thus, manipulation of a particular characteristic does not insure that it is a feature that is utilized in letter recognition (Cheng & Pachella, 1984). Which visual characteristics are features is a psychological question.

According to the model developed here, the wide variety of objects (features, letters, and words) in reading are recognized in accordance with a general pattern recognition algorithm (Massaro, 1979; Oden, 1979; Oden & Massaro, 1978). The model postulates three operations: featural evaluation, featural integration, and classification. Continuously valued features are evaluated, integrated with respect to prototype representations, and a classification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype. The model is called a fuzzy logical model of perception; the concept of fuzzy logic has been discussed by Goguen (1969), Oden (1977), and Zadeh (1965). How it is used in the model has been described by Oden and Massaro (1978) and Massaro and Cohen (1983a).

Featural evaluation assumes that the features functional in letter recognition are continuous rather than discrete. In addi-

Offprint requests to: D.W. Massaro

tion, the evaluation of each feature proceeds independently of the properties of the other features. Using the continuous truth values of fuzzy logic as a metric, featural evaluation is conceptualized as providing truth values, t(x), representing the degree to which each relevant feature is present. The features are defined by prototype descriptions of the letters which are stored in memory. All of the features of a letter are evaluated with respect to the degree to which the features support the various letter alternatives, such as a Q or a G.

Featural integration involves the integration of the truth values of the features with respect to the prototype representations. A prototype defines a percept or concept in terms of an arbitrarily complex fuzzy logical proposition. For example, a letter prototype would represent the conjunction of the visual features defining the letter. The integration operation consists of replacing the respective features of each prototype with their corresponding truth values from featural evaluation of the relevant test letter. The conjunction of these truth values determines to what degree each prototype is realized in the test letter. The outcome of featural integration consists of a goodness-of-match value for each prototype.

During pattern classification, the merit of each relevant prototype is evaluated relative to the summed merits of all relevant prototypes. The relative goodness of a prototype gives the proportion of times it is selected as a response or its judged magnitude. This decision process is similar to Luce's (1959) choice rule which is based on the relative strengths of the alternatives in the candidate set. In pandemonium-like terms (Selfridge, 1959), we might say that it is not how loud some demon is shouting but rather the relative loudness of that demon in the crowd of relevant demons. The likelihood of Q identification would be equal to the goodness-of-match values to the alternative Q relative to the sum of the goodness-of-match values for all of the relevant alternatives.

Factorial designs

Our method of study of perceptual recognition utilizes factorial designs that manipulate independently multiple characteristics of the letters. Consider, for example, how a range of letters between G and Q can be created when the obliqueness of a line and the openness of the gap in the letter Q are varied across seven levels each (Figure 1). Seven levels of openness are created by removing 0, 2, 3, 4, 7, 9, and 10 points from the oval of the capital letter Q. Similarly, the obliqueness of the line varies between the horizontal and 10.8, 20.9, 28.9, 37.9, 50.7, 60.9 degrees of obliqueness measured from the horizontal. The resultant 49 test letters are presented to subjects in randomized order and repeatedly for their identification as G or Q. Two dependent measures are the identification judgments and the reaction times (RTs) of these judgments. In addition to requiring discrete judgments, subjects are asked to rate each letter along a continuum between G and Q.

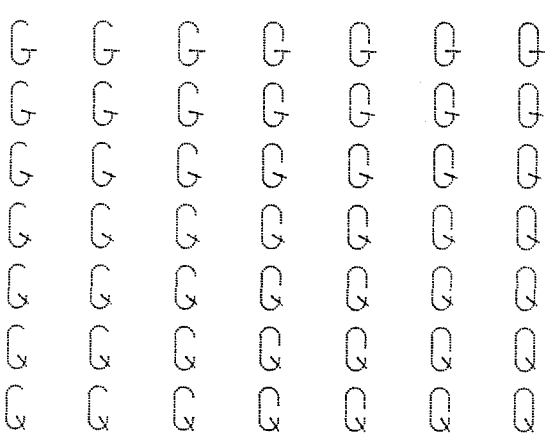

Fig. 1. The 49 Q/G test letters created by varying obliqueness of the line (row) and the openness of the gap in the oval (columns)

This paradigm can provide information about the nature of letter identification. We demonstrate that it can be used to address not only which visual characteristics are functional visual features, but also a variety of other questions, which are considered below. Before doing so, we present an overview of the experimental method used in the present studies.

General method

Stimuli were presented on a display screen (Tektronix 604) using a digital to analog converter. All experimental events and data collection were controlled by a PDP-11/34a computer.

The Q/G stimuli Figure 1 were presented to subjects by plotting the points of a test letter in a 80 by 240 grid on a display screen. When presented on the display, each test letter was about 0.22 inches wide and 0.47 high. Given a viewing distance of about 24 inches, the visual angle was 0.53 degrees for the width and 1.12 degrees for the height of the test letter. The test letters were plotted so that its center corresponded with the center of the display screen. The intensity of the display was set at a comfortable viewing level.

In all of the present experiments each trial began with the presentation of a fixation point for 500 ms on a display screen. A test letter immediately followed the fixation point. The length of the display was either 200 or 400 ms with the duration remaining constant within a particular experiment. After the stimulus presentation, the subjects made an identification or rating judgment of the letter. The next trial began 1 s after all subjects had responded. Up to four subjects could be tested simultaneously in separate sound attenuated rooms. The incandescent lighting in the rooms was dimmed to allow better viewing of the displays.

Subjects judged the test letters by using a discrete judgment or a continuous rating task. In the discrete judgment task, subjects were instructed to indicate whether the test letter presented on each trial was a Q or a G by pressing the appropriately labeled key on the computer terminal keyboard. For the rating task, subjects rated the "Q-ness"-"G-ness" of a test letter from the Q/G continuum by using a rating scale displayed on the computer terminal monitor. The scale was a straight horizontal line made up of 50 divisions. The left end of the scale was labeled "Q" and the right end "G". Subjects were able to move a pointer along the scale but were not told that the scale had 50 divisions. The pointer was represented as a black box on the rating scale and subjects manipulated the pointer using left and right arrow keys on the terminal keyboard. In the rating session subjects were instructed to "... tell us where the test letter falls on the scale from Q to G by moving the pointer on the screen in front of you ... we want you to use the whole Q-G scale, not just the two endpoints and middle, for example. For the letters you will see in this study, you should use the entire scale and all of the points in it." In both tasks, subjects were told that the test letters were presented in random order, with no patterns for them to guess.

In all tasks, subjects participated in a familiarization session and two experimental sessions lasting about 20–25 min with a five minute break between sessions. In the familiarization session, the Q/G stimuli were presented twice in random order so that the subjects could gain some sense of the range of differences among the test letters. Each experimental session began with 20 practice trials which were not included in the analyses.

Within each experimental session, the Q/G test letters were presented in 6 blocks of 49 experimental trials. Within each block, stimuli were sampled without replacement from the 49 levels of the Q/G continuum. Subjects completed 628 trials consisting of 40 practice judgments and 12 judgments for each of the 49 test letters across the two experimental sessions.

Three separate experiments were performed using the above method. In one experiment, nine subjects judged with the discrete choice task stimuli presented for 400 ms. In a second study we had three subjects use a discrete choice task to judge the test letters presented for 200 ms. Finally in a third experiment six subjects participated in the rating task in which the test letters were presented for 200 ms. We use these experiments to test the issues in perceptual recognition described in the next section.

Binary contrasts

Our research strategy follows the tenets of falsification and strong inference (Platt, 1964; Popper, 1959) in that binary oppositions are constructed and tested. In addition, we look for converging operations or converging results from a variety of experimental paradigms and behaviors (Garner, Hake, & Eriksen, 1956). The binary contrasts consist of alternative theoretical descriptions of psychological phenomena. The tests are constructed so that the results will be consistent with one theory and inconsistent with the other. Before the experiment, each theory has an equal opportunity to explain the results; the outcome of the experiments determines the winner. The dissection of psychological phenomena within the framework of binary oppositions, combined with the tools of information integration (Anderson, 1981, 1982) and mathematical-model testing, not only illuminates the phenomenon of letter recognition itself but also more general problems of perception and pattern recognition.

The explicit tests of fundamental questions also make apparent the implicit assumptions inherent in much of the current empirical and theoretical research. For example, most experimental tests assume that people process multiple features to make a perceptual judgment of a letter. This assumption bypasses a fundamental issue in visual pattern recognition: whether readers utilize all the featural information available or use only one critical feature when making their judgments (Oden, 1979; Massaro, 1979, 1985). By designing our experiments to address these issues, we not only examine how readers use visual information to recognize a letter but we also determine whether readers use more than one source of the available information when making perceptual judgments.

The binary oppositions to be considered in this paper are arranged hierarchically (Figure 2). In some cases, the question at one level is dependent on the answers to questions at higher levels. The template versus multiple features contrast determines the number of information sources available: a template model argues that a letter makes available only a single information source whereas a feature model argues for many sources of information. If readers have available multiple

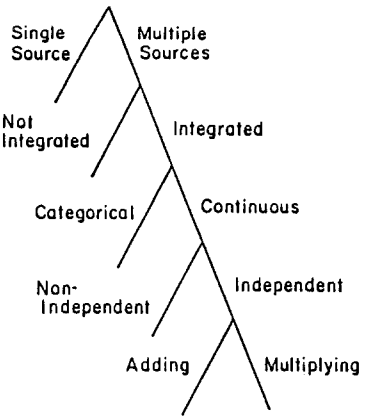

Fig. 2. An illustration of the binary oppositions central to the domain of letter recognition

sources of information about a given letter, then it is necessary to determine whether readers use one or many features out of the pool of available features when making perceptual judgments. We formulate this question in terms of integration versus nonintegration. The categorical versus continuous contrast determines whether the sources of information (e.g., features) are continuous or discrete. The nature of one source of information might depend on the nature of other sources and this issue is examined in terms of the dependence or independence of the sources of information. Finally, if features are integrated, we ask whether they are integrated in an additive or multiplicative fashion (Anderson, 1981, 1982). We now discuss the experimental tests of these issues.

Templates versus multiple features

The first question asks whether there are multiple sources or just a single source of information in letter recognition. By source of information we mean some property of the letter that is functional in letter recognition. This question is reminiscent of the traditional contrast between templates and features (Massaro & Schmuller, 1975; Neisser, 1967).

A template theory assumes that a reader uses a single source of information when identifying a letter pattern, that is the letter pattern itself. According to the template theory, a reader does not analyze a letter into component parts, but perceives it as a whole. To recognize the letter, a reader must compare the sensory experience of a test letter against templates, or psychological representations of letters in the perceptual system (Selfridge & Neisser, 1960; Gibson, 1963). The identity of a test letter is the name of the template that best matches the test letter. The template model, however, does not consider that the letter pattern might be represented to the perceptual system in terms of its constituent parts, as many studies on pattern recognition suggest (Palmer, 1975; Pomerantz & Garner, 1973).

An opposing view to the template approach is to assume that multiple sources of information are available, for example, a minimum set of common geometric components of letters, such as vertical bars and curved lines, that allow people to discriminate among all the letters of an alphabet (Bouma, 1971; Oden, 1979). To recognize a letter, a reader compares the letter features obtained from a physical letter pattern against the knowledge of what distinctive features represent the letter. Traditionally, distinctive features were assumed to

provide only binary information: the presence or absence of a physical subpattern of a character. Current thought assumes that features provide information about the degree to which they match a particular letter alternative (Naus & Shillman, 1976; Massaro, 1979; Oden, 1979).

We performed a qualitative test of the template versus multiple feature contrast. We had three subjects classify as Q or G the test letters from Figure 1 using the method described in the factorial section. Three subjects judged each of the 49 test letters, presented with a 200 ms duration, 12 times each. We calculated each subject's probability of responding Q for each test letter by dividing the number of times a test letter was identified as Q by 12.

According to the template model, changes in the probability of identifying a letter as Q should be based on the similarity of the shape of a test letter to a Q template. The template model is not sensitive to openness as a separate feature that distinguishes Q from other letters; for example, it just considers openness as part of the overall shape of the character. When we alter the shape of the test letter by changing openness, the template model predicts that the overall perceived similarity of the test letter to a Q template should change independently of the shape of the rest of the pattern. Therefore, changes in openness should have the same effect on the similarity of a test letter to a Q template regardless of the changes in the obliqueness of the line of the character. A test of this prediction would be to examine whether or not there is an interaction between the independent variables of obliqueness and openness. An interaction would indicate that changing the shape of one part of the test letter is not independent of changes in other part of the test letter. On the other hand, a lack of an interaction would be consistent with the template model.

Figure 3 plots the average probability of Q responses from the discrete choice experiment. The changes in the probability of responding Q appear to be based on changes in both the obliqueness and openness dimensions. In addition, the bow-shaped curves in Figure 3 indicate that obliqueness contributed more to Q judgment probabilities when openness was ambiguous and vice versa. An ANOVA with subjects, obliqueness and openness as factors revealed a significant interaction between obliqueness and openness, $F(36,72) = 15.47$, $P < 0.01$. The bow-shaped curves together with a significant interaction provides strong support for the idea that subjects use multiple sources of information when judging the test letter. The systematic change in the Q identification probabilities indicates that changes in one part of a test letter pattern change the influence of other parts of the pattern. Clearly, multiple sources of information are used in making the perceptual judgments.

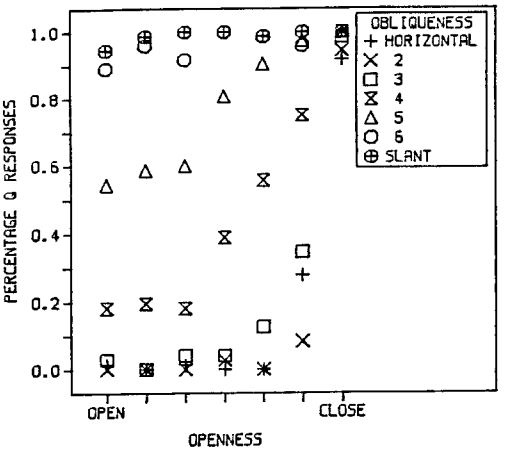

Fig. 3. Average percentage of Q identifications for three subjects as a function of openness and obliqueness

Integration versus nonintegration

Given multiple sources of information, an important question is whether these sources are integrated or combined in the letter recognition process. It is possible that multiple sources are functional but that only a single source is used during a given presentation of the letter.

The idea of only a single source of information being functional during a given presentation can be represented by a model that assumes that a subject can only evaluate one letter feature on any given trial in order to make a perceptual judgment. Consider the Q/G matrix illustrated in Figure 1 in which each test letter is formed by a factorial combination of the features of openness and obliqueness. Subjects identify these test letters as G and Q. The results allow us to test between the fuzzy logical model and the single feature model. The fuzzy logical model assumes that both the dimensions of obliqueness and openness are evaluated. The single feature model assumes that only one dimension or the other is evaluated given a test letter. We will now describe how the fuzzy logical and single feature models can fit the results of a Q/G identification experiment.

Fuzzy logical model. Given a test letter, the featural evaluation stage determines the degree to which the Q and G alternatives are supported by the visual information. Using fuzzy truth values, a value between zero and one is assigned to the obliqueness and openness dimensions, indicating the degree to which these features support the G and Q alternatives.

The features values of openness and obliqueness are then integrated by the Q and G prototypes. The prototypes are defined by:

Q: Not Open Oval & Oblique Line

G: Open Oval & Horizontal (Not Oblique) Line

Given a prototype's *independent* specifications for the obliqueness and openness features, the value of one feature cannot change the value of the other feature at the prototype matching stage. Using the definition of fuzzy negation as 1 minus the feature value (Oden, 1977) we can represent the prototypes in terms of openness and obliqueness.

Q: (1 – Openness) & Obliqueness

G: Openness & (1 – Obliqueness)

The integration of the features defining each prototype can be represented by the product of the feature values (Massaro & Oden, 1980; Oden, 1979; Oden & Massaro, 1978). In this case, the goodness of a Q or G alternative can be represented by:

$$G(Q) = (1 - t(\text{Openness})) \times t(\text{Obliqueness}) \quad (1)$$

$$G(G) = t(\text{Openness}) \times (1 - t(\text{Obliqueness})) \quad (2)$$

where G() represents the goodness of match of a test letter to the Q and G alternatives and t() is a function that determines the truth value of a particular feature: the degree to which a gap is open or a straight line is oblique.

If Q and G are the only valid response alternatives, the pattern classification operation determines their relative merit leading to the prediction:

$$P(Q) = G(Q)/(G(Q) + G(G)) \quad (3)$$

where P(Q) is the predicted probability of a Q response to a particular test letter shown in Figure 1.

Single feature model. This model assumes that the subject evaluates either the obliqueness or openness dimensions of a test letter but not both on a given trial. If a subject had to decide whether a test letter was a Q or a G, the subject would evaluate the obliqueness feature with probability w and the openness feature with probability l-w. Identification performance is pedicted to be a weighted average of these two kinds of trials. The probability that a subject would respond Q is:

$$P(Q) = w\,p(\text{oblique}) + (1-w)\,(1-p(\text{open})) \quad (4)$$

where w is the probability that a given feature will be selected on each trial, and p(oblique) and p(open) are the probabilities that a subject will respond Q when evaluating only the obliqueness or openness feature, respectively.

Nine subjects saw each test letter (Figure 1) for 400 ms 12 times in random order. On each trial they labeled the test letter as a Q or a G, and the probability of a Q response for each test letter was the dependent variable. Given that the Q and G identifications sum to one, P(Q) for each test letter completely represents the identification judgments. Thus we have 49 independent observations to describe the 49 test letters.

Both the single feature and the fuzzy logical model were fit to the data of the individual subjects using the computer program STEPIT (Chandler, 1969). A model is defined in STEPIT as prediction equations that contain a set of unknown parameters. STEPIT minimizes the deviations between the observed and predicted values of the models by iteratively adjusting the parameters of the equations. Root mean square deviation (RMSD) values determined the overall goodness of fit of the alternative models. This value is the square root of the mean of the squared deviations between the predicted and observed values. The smaller the RMSD value, the better the fit of the model.

Fourteen parameters are necessary to fit the fuzzy logical model to 49 data points: seven parameters for each level of obliqueness and openness. The parameters represent the degree to which the obliqueness and openness features match the Q alternative. Equations 1, 2, and 3 were then used to predict the probability of a Q response to a given letter.

Fifteen parameters were used to fit the single feature model, represented by Equation 4, to the data. One parameter estimates the probability that the subject would say Q for a given level of obliqueness or openness, resulting in a total of 14 parameters, one for each of the 7 levels of the 2 stimulus dimensions. One parameter estimates the probability that the oblique feature is selected on a given trial (see Equation 4).

Table 1 presents the RMSD values for each subject. The RMSDs for the single feature model range from 0.08 to 0.23.

Table 1. The root mean square deviations between observed and predicted values given by the models for the discrete judgment task

Subject	Single Feature	Fuzzy Logical
1	0.2347	0.0346
2	0.0820	0.0637
3	0.1422	0.0973
4	0.1983	0.0444
5	0.2102	0.0312
6	0.2241	0.0543
7	0.2290	0.0487
8	0.2142	0.0284
9	0.1917	0.0369

In contrast, the RMSDs for the fuzzy logical model range from 0.03 to 0.10. Thus, the quantitative model tests of the individual subject data indicate that the multiple feature model did much better in accounting for the variance in the data.

The average predicted versus observed Q probabilities for the fuzzy logical and the single feature model are illustrated in Figures 4 and 5. The lines in the figure represent the predicted values. Figures 4 clearly shows the poor match between the observed data and the predictions given by the single feature model. The predictions cannot capture the statistical interaction between the two independent variables. In contrast, the predictions of the fuzzy logical model give a much better match to the observed data, as seen in Figure 5. The shape of the curves in the prediction lines do very well in capturing the trends in the data.

Evidence for the integration of featural information in letter perception comes from fit of the fuzzy logical and single feature models. The single feature model assumes that subject use only one information source in order to make a particular perceptual judgment. However, the superior fit of the fuzzy logical model indicates that subjects not only use multiple sources of information, but they also integrate this information from multiple sources every time they make a perceptual judgment.

Continuous versus categorical perception

An important issue in letter recognition is whether perception is categorical or continuous. Given that multiple features are integrated in perceptual recognition, one contrast is to consider whether featural information is continuous or categorical. Categorical information implies that the perceptual system categorizes a feature such as obliqueness as either representative or not of the Q or G alternative before integration occurs. On the other hand, continuous information implies that the perceptual system evaluates and maintains the degree to which each feature matches Q or G for the integration process. Massaro and Cohen (1983b) formulated a categorical model and tested it against the continuous fuzzy logical model. The categorical model is mathematically equivalent to the single feature model formulated and tested in the previous section. It follows that a rejection of the single feature model also permits rejection of the categorical model and we can conclude that the featural information is continuous at the time of integration.

A second form of categorical perception requires consideration. In this model, the outcome of the integration process is categorical even though the featural information might have been continuous. That is, the information available for any judgment about letter identity is either categorical or continuous. A discriminating test at this level of the contrast is possible by an analysis of the distribution of rating responses to repeated presentations of a stimulus event (Massaro & Cohen, 1983a). Oden (1979) provided convincing evidence for continuous perception of letters in his original study of varying multiple features in a factorial design. The single judgments of a given test letter could not be described by a categorical model but were well described by the same continuous model developed here.

Consider the Q/G continuum presented in Figure 1. Categorical perception predicts that the ratings to repeated presentations of a test letter will come from two kinds of trials: those trials on which the test letter was identified as Q and those on

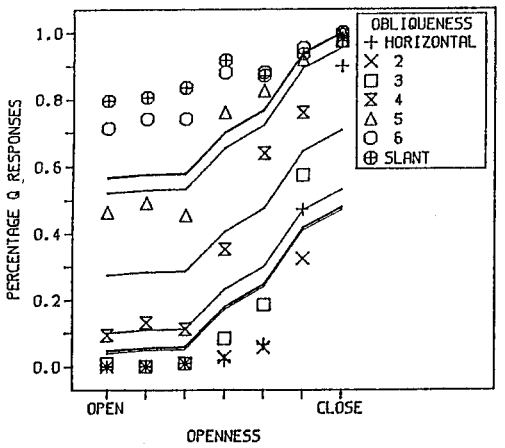

Fig. 4. Average percentage of Q identifications for nine subjects as a function of openness and obliqueness. The lines in the figure give the average predictions of the single feature model

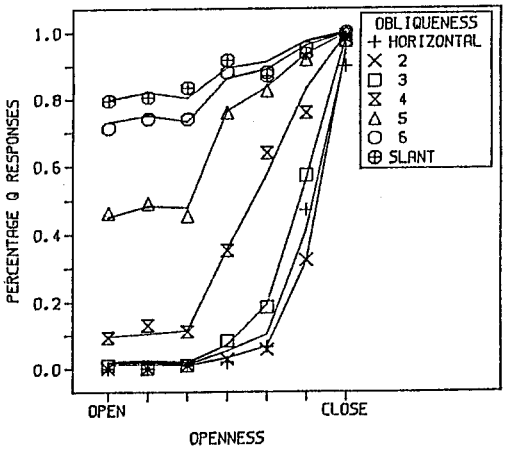

Fig. 5. Average percentage of Q identifications for nine subjects as a function of openness and obliqueness. The lines in the figure give the average predictions of the fuzzy logical model

which the test letter was identified as G. Thus, categorical perception predicts that the distributions of ratings to a given test letter is a result of the two different Q/G letter categorizations. If the categorical model accurately describes rating performance, then rating judgments should produce a bimodal distribution of ratings across the rating scale, one peak for each of the two letter categories. On the other hand, continuous perception predicts that the rating is based on continuous information representing the degree to which the test letter matches the letter alternatives. Hence, the distribution of ratings to a given test letter will result from a single kind of trial on which the perceiver has continuous information about the test letters, resulting in a unimodal rating distribution.

To examine this prediction, six subjects rated the degree of Q-ness or G-ness in the letters in Figure 1. Each test letter was displayed for 200 ms and was rated 12 times each by six subjects. Figure 6 presents the distributions of ratings for a typical subject for a 3 × 3 subset of the 49 test letters of Figure 1. As can be seen in the figure, it is very difficult to see how these ratings could have resulted from a mixture of two different distributions; the ratings represent a unimodal distribution for each test letter. Supporting this observation, mathematical models embodying the categorical and continuous assumptions were formalized and fit to the distributions of ratings. The continuous model gave a better description than the categorical model for each of the six subjects, even though the categorical model required almost twice as many free parameters. Thus, we have evidence based on the distribution of ratings that the test letter is perceived continuously rather than categorically.

A second test of the categorical/continuous contrast at the level of letter perception is to examine identification reaction times. The continuous model predicts that the reaction time for identification would depend on the degree to which the integrated featural information provides unambiguous support for a given letter category. Continuous information will vary in the degree to which the test letter is representative of a given letter alternative. Ambiguous information should increase identification reaction times relative to unambiguous featural information. Categorical perception of a letter either leads to a Q or a G without any index of ambiguity. The time that a given decision takes should remain the same regardless of the nature of the test letter.

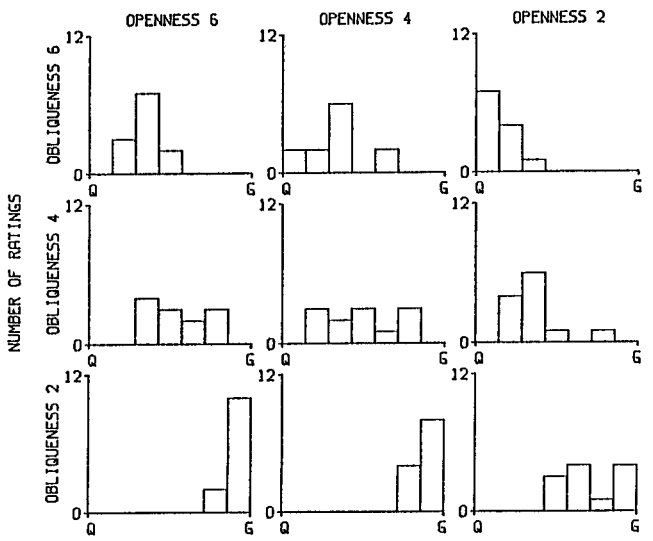

Fig. 6. Frequency distributions of ratings between Q and G of a typical subject to a 3 × 3 subset of the 49 test letters of Fig. 1. Obliqueness ranges from horizontal (level 1) to slant (level 7) and openness ranges from close (level 1) to open (level 7). The categorical model predicts that repeated ratings to a given test letter result from two distributions whereas the continuous model predicts that the ratings result from a single distribution

We evaluated reaction times as a function of ambiguity to assess whether letter perception is continuous or categorical. We used the data from the experiment in which three subjects identified the 49 test letters (Figure 1) as Q or G. We then assumed that test letters judged as Q 50% of the time were the most ambiguous and the test letters judged as Q or G 100% of the time were the least ambiguous. An ambiguity value was computed by taking the absolute difference between P(Q) and 0.5 and subtracting it from 1:

$$A = 1 - |P(Q) - 0.5| \quad (5)$$

where P(Q) is the probability of a Q response. Figure 7 presents a scatter plot of reaction times as a function of this ambiguity for the 49 Q/G test letters. The correlation between ambiguity and reaction time accounts for about 49% of the variance in the data. Therefore, there is good evidence that

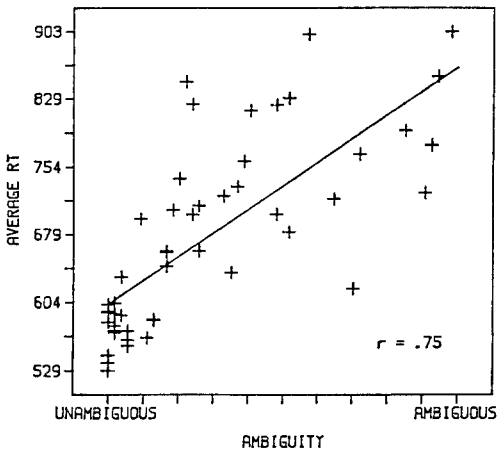

Fig. 7. A scatter plot of average reaction times of 3 subjects as a function of ambiguity for the discrete judgments of the 49 test letters. The solid line represents the regression line for the correlation

reaction times increase with increases in ambiguity as measured by the likelihood of a given response. This result rejects the hypothesis of categorical letter perception and supports the claim that readers have continuous information about the test letter.

The results of three types of studies provide converging evidence for continuous perception. Model tests of identification judgments give evidence for continuous featural information. The unimodal distribution of rating judgments indicates that subjects did not transform the visual information of the test character into a discrete category prior to their rating judgments. Finally, reaction times reveal that a given discrete decision is sensitive to the nature of the letter information leading to the decision. In light of the evidence for continuous perception from the rating task, we can conclude that the discrete judgment task required a categorical decision to a continuous percept. Therefore it is not surprising that a subject would require additional time to decide whether a test letter was a Q or a G to the extent that the visual information was ambiguous. In terms of signal detection theory, reaction times would increase with decreases between the test letter observation and some criterion separating the Q and C categories along the Q/G perceptual dimension.

Independent versus dependent evaluation of information sources

The fourth contrast has to do with the independence of the multiple sources of information in letter recognition. Are the sources of information treated independently in the letter recognition process or is each source colored by the nature of the other sources that occur with it?

One way of testing this contrast is to ask whether a model assuming independent evaluation of features provides an adequate account of letter perception. The hypothesis of dependence must predict a failure of any model assuming independence. Independence models assume that the information obtained along one source is independent of the information obtained along the other source. One such model is the fuzzy logical model. If the obliqueness and openness features were dependent, the changes in openness would influence the amount of information transmitted by obliqueness. In contrast the fuzzy logical model assumes that these parameters are free to vary independently of each other. Since the fuzzy logical model can capture most of the variance in the data, as shown in a previous section, it seems nonparsimonious to assume that letter features are dependent. Indeed, as demonstrated in other domains (Oden, 1981; Massaro & Oden, 1980) the assumption of independence holds up well under many types of perceptual judgments. Therefore, it seems safe to conclude that features are treated independently.

Another test that might be done is to compare the identification reaction times of single features that compose the test letters to the identification reaction times of the test letters themselves. For example, subjects might be presented with the seven oblique line levels for Figure 1 and be asked to identify the lines as features consistent with the alternatives Q or G. Alternatively, subjects might judge openness of the oval presented in isolation in the same fashion. The reaction times to the single features could then be compared to the reaction times of the features presented together. If the features are not independent, then it should not be possible to account for the reaction times to a combination of the openness and obliqueness features in terms of the reaction times to openness and obliqueness features presented alone. If the two dimensions are independent, we might expect reaction times to a combination of features to be somewhat faster than those to the single features, but the advantage should be completely accounted for by statistical facilitation (Gielen, Schmidt, & Van Den Heuvel, 1983; Raab, 1962).

Additive versus multiplicative integration

At this point in our research tree, we have good evidence that letter perception involves integrating multiple sources of independent featural information. It is now important to determine the nature of featural integration. Both additive and multiplicative integration rules are promising models of the featural integration stage. The additive rule (Anderson, 1981, 1982) makes strong predictions about the average rating response in an integration task; if a subject rates a test letter that varies on two factors on an interval scale, then the plot of the ratings versus the factors should produce parallel lines. The additive rule assumes that the effect of each factor on perceptual integration is the same regardless of the ambiguity of other factors. This rule is not optimal in that averaging an ambiguous source of information with an informative source will tend to neutralize the judgment relative to the informative source presented alone. In contrast, the multiplicative rule predicts American football-shaped curves when the average ratings are plotted in a two factor graph. The curves indicate that the least ambiguous source of information has the most impact in perceptual judgments.

A test of the additive/multiplicative contrast was performed by utilizing the results of the rating experiment described above. The effectiveness of the models in accounting for the data can be seen by fitting additive and multiplicative models of the integration process to the rating results using the program STEPIT (Oden, 1979; Massaro, 1979). The multiplicative model of integration for the Q/G continuum is represented by Equations 1 and 2. Additive integration is given by:

$$G(Q) = (1 - t(\text{openness})) + t(\text{obliqueness}) \quad (6)$$

$$G(G) = t(\text{openness}) + (1 - t(\text{obliqueness})) \quad (7)$$

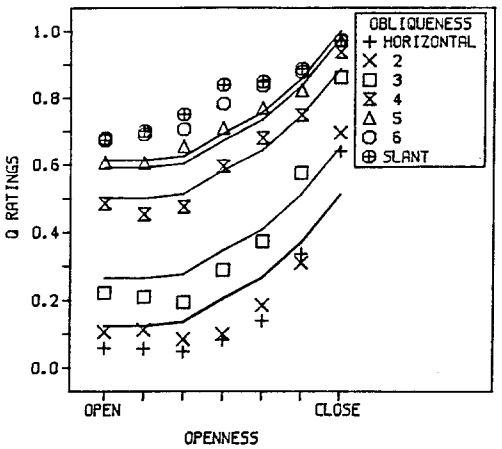

Fig. 8. Average Q ratings as a function of openness and obliqueness. The lines in the figure give the predictions of an additive integration rule

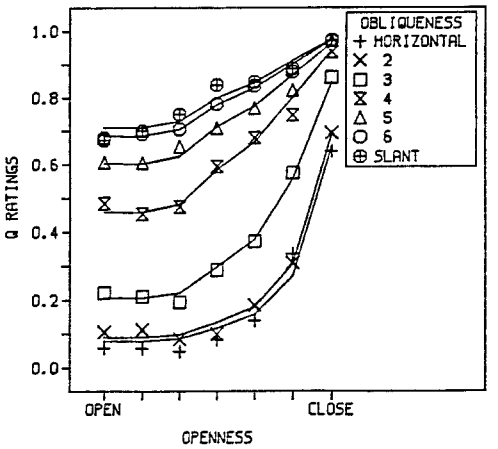

Fig. 9. Average Q ratings as a function of openness and obliqueness. The lines in the figure give the predictions of a multiplicative integration rule

Table 2. The root mean square deviations between observed and predicted values given by the models for the rating tasks

Subject	Additive	Multiplicative
1	0.0713	0.0363
2	0.1639	0.0365
3	0.0650	0.0507
4	0.1036	0.0321
5	0.0834	0.0420
6	0.0502	0.0335

We fit both models with 14 parameters, 1 parameter for each level of the two features. The relative goodness rule given by the pattern classification operation (Equation 3) was then used to predict the rating judgments for both the additive and multiplicative models for each subject. The solid lines in Figures 8 and 9 illustrate the additive and multiplicative model fits averaged over subjects, respectively. The additive model (Figure 8) predicts parallel lines which do a rather poor job in fitting the data points. The multiplicative model (Figure 9) with its bowed shaped predictions, does much better than the additive model. The RMSDs for the individual subjects are presented in Table 2. The RMSDs support the conclusion drawn from the figures: that the multiplicative model fits the individual subjects' ratings much better than the additive model.

The results provide insight about the operation of perceptual integration. The multiplicative model's superiority over the additive model suggests that perceptual integration acts in an optimal manner. The least ambiguous source of information will contribute the most to the rating judgment.

Additional findings

In addition to addressing fundamental issues in letter recognition, the present framework provides information about specific properties of the processes involved. Consider what appears to be an asymmetry in the results shown in Figure 10. These results are identical to those in Figure 5, but the plot of the two independent variables has been interchanged to make the asymmetry more apparent. This asymmetry is most evident for the top two lines in the figure. When the oval is closed (openness level 7), the probability of a Q judgment is nearly 100% regardless of the obliqueness of the line. When the oval is open by only a small amount (openness level 6) the angle of the line has a very strong effect on the probability of responding Q. The asymmetry caused by oval closure might be explained by the poor quality-of-Xeroxing effect. The closed oval influences the reader's judgment much more than an open oval since poor copying (or poor vision) could have been responsible for the absence. In contrast, the closed oval is unlikely to result from poor copying (or poor vision). Thus, the reader will give more weight to the presence than the absence of the oval. In this case, weight corresponds to the feature value; presence of the oval will be more extreme than absence. Supporting this analysis, the parameter value for least openness (level 7) is more extreme than for the most openness (level 1). Since oval presence is characteristic of a Q prototype and not a G prototype (see Equations 1 and 2), a reader is more likely to see a closed oval character as Q regardless of the obliqueness of the line.

Extensions

Although we stressed the role of letter recognition in reading, we are capable of analyzing higher-order contributions within the framework developed here. The integration of continuous bottom-up sources of information in letter recognition is easily extended to include contributions from higher-order context (Massaro, 1979; Oden, 1984). Each component of higher-order context is treated simply as an additional source of information contributing to perceptual recognition. The results of a variety of studies are consistent with the independence of bottom-up and top-down sources. Some top-down source, such as lexical constraints does not modify the bottom-up sources; it simply adds an independent and continuous source of information to be integrated with the bottom-up source. The integration of a top-down source and bottom-up source seems to follow the same multiplicative rule given by the fuzzy logical model for integrating two bottom-up sources (Massaro, 1979; Oden, 1984). The binary set of contrasts in Figure 2 can be read in terms of bottom-up and top-down sources. In this case, as in the case of visual features, the outcomes of the contrast go down the right side of the tree.

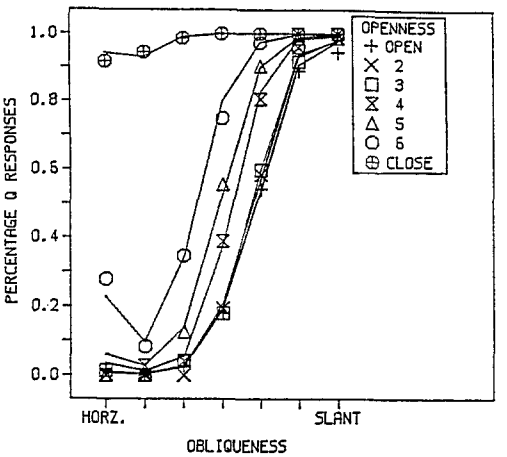

Fig. 10. Average percentage of Q identifications for nine subjects as a function of openness and obliqueness. The lines in the figure give the average predictions of the fuzzy logical model

Conclusion

We have approached the problem of letter perception within the framework of falsification and strong inference. The binary contrasts have been successful in eliminating plausible and intuitive interpretations of letter perception. The methods of information integration and mathematical model testing appear to be ideally suited for addressing the issues. The results also illuminate many aspects of the psychological processes involved. Readers appear to evaluate multiple letter features when making perceptual judgments. The evaluation of one feature seems to occur independently of the properties of the other features. This evaluation process makes available continuous information indicating the degree to which relevant alternatives are supported. The integration process is not simply a compromising operation in that only mild support from each of the two features can be integrated to produce strong support for a letter alternative.

The issues that we have addressed seem fundamental to developing a psychological understanding of letter recognition. Perceiving letters is described within the context of a general theory of perceptual recognition. This theory provides a common metric for evaluating and integrating multiple sources of information in pattern classification. Future work will be necessary in order to explore variations within the context of each binary contrast. We can also expect other contrasts and theoretical alternatives to present themselves as our understanding of letter perception evolves.

References

Anderson, N.H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N.H. (1982). *Methods of information integration theory.* New York: Academic Press.

Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research, 11,* 459–474.

Carrithers, C., & Bever, T.G.R. (1984). Eye-fixation patterns during reading confirms theories of language comprehension. *Cognitive Science, 8,* 157–172.

Cattell, J. McK. (1886). The time it takes to see and name objects. *Mind, 11,* 63–65.

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

Cheng, P.W., & Pachella, R.G. (1984). A psychophysical approach to dimensional separability. *Cognitive Psychology, 16,* 279–304.

Garner, W.R., Hake, H.W., & Eriksen, C.W. (1956). Operationism and the concept of perception. *Psychological Review, 63,* 149–159.

Gibson, E.J. (1963). Perceptual learning. *Annual Review of Psychology, 14,* 29–56.

Gielen, S.C., Schmidt, R.A., & Van Den Heuvel, P.J.M. (1983). On the nature of intersensory facilitation of reaction time. *Perception & Psychophysics, 34,* 161–168.

Gilmore, G.C., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics, 25,* 425–431.

Gough, P.B. (1984). Word recognition. In Pearson, P.D., Barr, R., Kamil, M.L., & Mosenthal, P. (Eds.), *Handbook of Reading Research (pp. 225–253).* New York: Longman.

Goguen, J.A. (1969). The logic of inexact concepts. *Synthese, 19,* 325–373.

Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354.

Loomis, J.M. (1982). Analysis of tactile and visual confusion matrices. *Perception & Psychophysics, 31,* 41–52.

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. (1984). Building and testing models of reading processes. In P.D. Pearson, R. Barr, M.L. Kamil, & P. Mosenthal (Eds.), *Handbook of Reading Research* (pp. 225–253). New York: Longman.

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

Massaro, D.W., & Cohen, M.M. (1983a). Categorical or continuous speech perception: A new test. *Speech Communication, 2,* 15–35.

Massaro, D.W., & Cohen, M.M. (1983b). 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., & Oden, G.C. (1980). Speech perception: A framework for research and theory. In N.J. Lass (Ed.), *Speech and Language: Vol. 3. Advances in Basic Research and Practice,* New York: Academic Press.

Massaro, D.W., & Schmuller, J. (1975). Visual features, preperceptual storage, and processing time in reading. In D.W. Massaro (Ed.), *Understanding Language* (pp. 207–240) New York: Academic Press.

Naus, M.J., & Shillman, R.J. (1976). Why a Y is not a V: A new look at the distinctive features of letters. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 394–400.

Neisser, U. (1967). *Cognitive Psychology.* New York: Appleton-Century-Crofts.

Oden, G.C. (1977). Integration of fuzzy logical information, *Journal of Experimental Psychology: Human Perception and Performance, 3,* 565–575.

Oden, G.C. (1979). A fuzzy logical model of letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 336–352.

Oden, G.C. (1981). Fuzzy propositional approach to psycholinguistic problems: An application of fuzzy set theory in cognitive science. In M.M. Gupta, R.K. Ragade, & R.R. Yager (Eds.), *Advances in Fuzzy Set Theory and Applications.* Amsterdam: North-Holland.

Oden, G.C. (1984). Dependence, independence and emergence of word features. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 394–405.

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

Paap, K.L., Newsome, S.L., & Noel, R.W. (1984). Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 413–428.

Palmer, S.E. (1975). Visual perception and world knowledge. In D.A. Norman & D.E. Rumelhart (Eds.) *Explorations in cognition.* San Francisco: Freeman.

Palmer, S.E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and Categorization,* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Perfetti, C.A., Goldman, S.R., & Hogaboam, T.W. (1979). Reading skill and the identification of words in connected discourse. *Memory & Cognition, 7,* 273–282.

Platt, J.R. (1964). Strong inference. *Science, 146,* 347–353.

Pomerantz, J.R., & Garner, W.R. (1973).Stimulus configuration in selective attention tasks. *Perception & Psychophysics, 14,* 565–569.

Popper, K. (1959). *The Logic of Scientific Discovery.* New York: Basic Books.

Raab, D.H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24,* 574–590.

Selfridge, O.G. (1959). Pandemonium: A paradigm for learning. In *Symposium on the mechanisms of thought processes.* London: HM Stationary Office.

Selfridge, O.G., & Neisser, U. (1960, August). Pattern recognition by machine. *Scientific American, 203,* 60–68.

Townsend, J.J. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics, 9,* 40–50.

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

Received May 14, 1986/August 10, 1986