HANDBOOK OF

Reading Research

EDITOR

P. David Pearson

SECTION EDITORS

Rebecca Barr

Michael L. Kamil

Peter Mosenthal

Longman

New York & London

Wilcox, K. Ethnography as a methodology and its applications to the study of schooling: A review. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational anthropology in action*. New York: Holt, Rinehart & Winston, 1982.
Wilkinson, L. C. (Ed.). *Communicating in the classroom*. New York: Academic Press, 1982.

Dominic W. Massaro
Perceptual Processing and
Recognition of Speech

5 BUILDING AND TESTING MODELS OF READING PROCESSES

EXAMPLES FROM WORD RECOGNITION

Dominic W. Massaro

Theories are nets cast to catch what we call "the world": to rationalize, to explain, and to master it. We endeavor to make the mesh ever finer and finer.

Karl R. Popper, 1959, p. 59

The measure of success in moving toward scientific explanation is the degree to which a theory brings out relationships between otherwise distinct and independent clusters of phenomena.

William K. Estes, 1979, p. 47

In the field of empirical sciences, . . . [the scientist] constructs hypotheses, or systems of theories, and tests them against experience by observation and experiment.

Karl R. Popper, 1959, p. 27

A theory is only overthrown by a better theory, never merely by contradictory facts.

J. B. Conant, 1947, p. 36

. . . To completely analyze what we do when we read would almost be the acme of a psychologist's achievements. . . .

E. B. Huey, 1908, p. 6

Building and testing models of reading processes advance our understanding of what the reader does while reading. We seek to know not only the nature of reading behavior but, more importantly, the internal mental machinery guiding the observable behavior. In this chapter, I discuss the logic of scientific endeavor in reading-related research. In addition, I present specific instances to

Source: The writing of this paper was supported, in part, by funds from the National Institute of Education and the National Institute of Mental Health. I appreciate the comments of Lola L. Lopes and Peter Mosenthal on an earlier version of this chapter.

illustrate some of the methods and techniques involved in the art of model building and testing. Consonant with my research interests and experience, the focus of specific models discussed here is word recognition in reading. The discussion is illustrative of model building and testing in general, however, and ought to apply to other aspects of reading conduct.

Although students of reading are fundamentally concerned with actual reading, seldom is an empirical or theoretical study directly concerned with normal reading. As is common in most areas of scientific endeavor, investigators feel it necessary to abstract for study only one or two components from the phenomenon of interest. The phenomenon of reading is complex in its entirety, and usually little knowledge is gained without some simplification and control in its investigation. Although some influential researchers (e.g., Neisser, 1976) are unhappy with the analysis of specific experimental situations, the most advanced state of the art in reading rests on exactly this kind of analysis (Baron, 1978; Estes, 1978; Gibson & Levin, 1975; LaBerge & Samuels, 1977; Massaro, 1975a, 1978, 1979b).

Given this state of affairs, we are concerned here with models of specific processes that occur in reading, rather than with general viewpoints of normal reading. These models of reading processes are usually developed and tested within the framework of laboratory experiments, and their exact relationship to normal reading is necessarily indirect. I do not mean to free the scientist from justifying a relationship between the simpler laboratory situation and the more complex act of normal reading. External validity, or the generalization from experiment to application, must be of constant concern. As an example, tachistoscopic studies reveal that visual perception during a given eye fixation is limited to a small area around the fixation point (Woodworth, 1938). That this limitation also exists in normal reading has been verified by manipulating the amount of text in view during any fixation in reading continuous text (McConkie & Rayner, 1975). Past experience in both the physical and psychological sciences supports the laboratory approach, although indirect, as the best choice for the researcher and theorist seeking an understanding of what we do when we read.

PSYCHOLOGICAL VERSUS PHYSIOLOGICAL MODELS

We are concerned with model building and testing in the study of psychological processes fundamental to reading. Models in this domain aim at a psychological rather than physiological level of description. Although a psychological model may one day be reduced to a physiological level of description, the psychological level will probably continue to be valuable. A good model enhances our understanding and leads to productive advances in research and practice. A psychological model may be easier to apply to the study of reading because its level of description may be more appropriate for assessment, intervention, and control.

The student may question whether the scientist can achieve a model of psychological processes that is independent of an understanding of the underlying physiological hardware. Consider the arrival on our planet of intelligent machines from outer space, intrigued by these general-purpose computing devices called people. The machines have no knowledge of, or experience with, living organisms. Our visitors could learn a great deal about how we function without addressing the unique properties of living matter. Their models of our behavior would resemble psychological models. The basic components of the models would involve hypothetical structures and processes that would relate certain observable behaviors to certain environmental situations. We do not doubt that these intelligent machines could develop something of truth in their psychological models. People are relatively good at understanding other people, even without great insights into underlying physiological mechanisms.

A psychological model is analogous to a model of a computer's software programs rather than its physical hardware. A computer can be built with vacuum tubes, magnetic chips, or large-scale integrated circuits. The underlying physical components are largely irrelevant to what programs can or cannot be run on the machine. Therefore, the potential user of the machine benefits more from learning about the machine's programs and how they are implemented than from studying its physical hardware. In some instances the machine's hardware constrains certain programming operations. This is analogous to the limitations in our visual acuity in reading, given the physiological makeup of the visual system. On the other hand, it is not apparent how any knowledge of underlying physiology can address the precise role of speech recoding in reading. Given the current state of the art, psychological models can provide the most valuable insights into reading processes.

SCIENTIFIC FRAMEWORKS

Models are usually developed and tested within general scientific frameworks, whether these frameworks are explicitly defined or only implicitly assumed. One framework for scientific endeavor has been expressed most succinctly by Popper (1959). The central assumption is that model building and testing must follow deductive rather than inductive methods. Following Hume, Popper claims that we are not justified in inferring universal statements from singular ones. Any conclusion drawn inductively may turn out to be false. Although we may have many instances of bright graduate students, this does not justify the conclusion that all graduate students are bright. Therefore the scientist should not try to verify a theory by demonstrating that it works in specific instances. Since new instances can always lead to the rejection of the theory, no experimental observation can verify a theory.

Popper proposes that theories,[1] once constructed, must be subjected to the following analyses: The investigator begins by comparing the conclusions derived from the theory in order to determine whether they are internally consistent. An analysis of the conclusions also indicates whether the theory is testable. By contrasting this theory with other theories, the investigator then determines whether the theory is unique and whether it would constitute a scientific advance if it survived experimental tests. Finally, if the conclusions drawn from the theory meet these requirements, then it is worthwhile to test the theory by subjecting its conclusions to experimental tests.

Experimental tests decide how well the theory survives. If it survives the experimental tests, we should not discard the theory. If experimental tests falsify conclusions drawn from the theory, then the theory should be rejected or modified accordingly. Surviving a particular experimental test only temporarily supports the theory because another investigator may provide a test that overthrows it. A critical feature of Popper's metatheoretical framework for building models is that verifiability and falsifiability of models do not have a symmetrical relationship. Although models can be falsified, they cannot be truly verified. Popper proposes that it is best to conclude that positive results only corroborate a particular theory; they do not verify it.

In a slightly different approach to scientific endeavor, Platt (1964) encourages scientists to employ a strong inference strategy of testing hypotheses. In contrast to generating a particular model, Platt would have the scientist generate multiple hypotheses relevant to a phenomenon of interest. The experimental test would be designed to eliminate (or in Popper's words, falsify) as many hypotheses as possible. The results of the experimentation would allow the generation of new hypotheses, which could be subjected to further tests. Both Platt and Popper adhere to David Hume's axiom prohibiting inductive arguments. The message is that the scientist should not attempt to confirm a single pet hypothesis. Platt's solution seems more productive, however, in that at least one of the multiple hypotheses under test should fail, and therefore the corresponding model can be rejected.

Given some ideal of model building and testing in empirical science, it becomes important to determine our natural disposition in formulating and testing hypotheses. An ingenious experiment carried out by Mynatt, Doherty, and Tweney (1977) was aimed at exactly this question. These researchers created an artificial research environment and allowed college subjects to make preliminary alterations in the environment. After observing the environment, the participants were instructed to formulate a hypothesis to account for the behavior of objects in the environment. The subjects were then given the opportunity to test their hypotheses. One group of subjects was instructed that the job of the scientist was to confirm hypotheses. Another group was told that the job was to disconfirm hypotheses. A third group was given neutral instructions in terms of simply testing hypotheses. Subjects were shown hypothetical environments that would allow tests of their hypotheses. Given a pair of environments, a subject was asked to choose one member of the pair to test the hypothesis. By evaluating these choices against the subject's original hypothesis, the authors determined whether the subjects chose situations to confirm or disconfirm their hypotheses.

Subjects chose a situation to confirm their hypothesis about seven times out of ten, regardless of how they were instructed. These results and other evidence adduced by the authors revealed a strong confirmation bias in this simulated research inquiry. People chose hypothetical situations to confirm their hypothesis and avoided consideration and testing of alternative hypotheses. A second result revealed that subjects faced with disconfirming evidence changed to a new hypothesis. Negative results were usually effective in changing the participant's opinion of the truth of the hypothesis. Although subjects may not seek disconfirmatory evidence, they use it correctly when it is available.

In a second study by the same authors (Mynatt, Doherty, & Tweney, 1978), subjects were faced with a much more complex research environment. Subjects attempted to determine how 27 fixed objects influenced the direction of a moving

particle in a two-dimensional display. The objects differed in size, shape, and brightness, and their influence on the particle varied as a function of these dimensions. A confirmation bias was again observed and could not be modified with highly explicit training in strong inference. In contrast to the previous study, subjects often kept or returned to their disconfirmed hypotheses. The more complex environment seemed to limit the generation of new hypotheses, and subjects therefore kept alive the few ideas they had. This is reminiscent of the compulsive gambler's dilemma: "I know the game is crooked, but it's the only game in town." A final result indicated that complete abandonment of a disconfirmed hypothesis also proved unproductive. For example, a general hypothesis of how objects influence particle direction was abandoned when it failed experimental tests. Nevertheless, the principle itself was correct, but only for some of the objects. Accordingly, it can be worthwhile to modify disconfirmed hypotheses based on previous results rather than reject all aspects of the disconfirmed hypothesis.

Continuing their research, Doherty, Mynatt, Tweney, and Schiano (1979) had subjects decide from which of two islands an archeological find had come. Subjects usually asked for information relevant only to their preferred island, not for information relevant to both islands. The requested information was useless because knowing that a characteristic of the found object is representative of one island does not provide information about whether the characteristic is also representative of the other island. This confirmation bias led to an inappropriate confidence in the subject's hypothesis. For example, if the subject believes that the pot with a curved handle comes from one island and then finds out that 80 percent of the pots from that island have curved handles, he or she becomes even more convinced that the pot is from that island. What the subject fails to realize is that pots from the other island may be even more likely to have curved handles.

Although the behavior of subjects who are not scientists in a simulated research environment cannot be generalized to empirical science, the results are illuminating and thought provoking. Interviews with NASA scientists, in fact, reveal a strong confirmation bias (Mitroff, 1974). A scientist tends to view the world in terms of a personal theory; he or she may be the least-equipped person to invent a critical test of the theory. Empirical tests usually turn out to provide results consistent with the theory without really providing a critical test of the theory. Given the conditional statement "If p, then q," where p is a statement concerning one side of a card and q a statement concerning the other side, and the alternatives are p, q, not-p, and not-q, what cards would you turn over to test the hypothesis? Most people choose only "p" or "p and q," when in fact "p and not-q" are the only critical tests of the hypothesis.

Most of the theoretically driven research in reading aims at supporting, not falsifying, theories. Given a natural predisposition for the scientist to verify rather than falsify a pet theory, it is important that the scientific endeavor remain competitive. The competitive dimension is probably as fundamental to the game of scientific endeavor as it is in most games (McCain & Segal, 1973). Scientists are much more willing to falsify other theories while seeking support for their own. If one accepts the ideal of falsifying models, the scientists are usually making most progress when they test someone else's theory, not their own. A model is a good model to the extent that its conclusions are easily tested by other scientists. Experimental tests should not only allow falsification of the model but also should provide clues to building new models if falsification occurs. Of course, these new models must also be easily testable.

Popper's and Platt's framework may be most appropriate for a relatively mature scientific discipline. In an early stage of model development, an empirical scientist may have to be content with a relatively facile acceptance of models or systems from other domains. As an example, Descartes used the machine as an analogue of overt behavior of people; the behavior of a person was viewed as following the same principles as those for machines. Models derived from analogues can be tested to some extent in that conclusions drawn from the analogue can be tested in the system of interest. Using the pump as an analogue for a model of the circulatory system, it is possible to derive certain relationships between pressure and output flow in pumps and test whether this relationship holds in the circulatory system.

Roediger (1980) discusses analogies used for models of memory. As examples, a house, purse, leaky bucket, dictionary, and garbage can serve as analogues of memory processes. A familiar example in cognitive psychology is that of the computing device as a model for some human mental functioning. The computing device can be described in terms of a number of stages beginning with input to the computer, operations on this input with respect to information stored in the memory of the computer, and output of the outcomes of the operations. In order to test whether the computer provides a good analogue to mental functioning, a more specific analogue is necessary to generate testable hypotheses. As an example, one could develop computer programs that carry out specific kinds of memory search and test if any of them describe how humans carry out memory search.

In some situations, an analogue can be shown to produce the same outcome as a human, but with an entirely different set of processes. If this is the case, the analogue is a poor model, even though it makes correct predictions. Work in artificial intelligence has encouraged psychologists to adopt working systems in this area as models of psychological processing. For example, speech-understanding systems have been used to guide the development of a model of reading (Allport, 1977; Rumelhart, 1977). One problem with this artificial system as a model for psychological functioning is that its computing capacity and speed can compensate for the deficiency of certain intelligent processes by sheer processing energy and time. Although an artificial system may be shown to perform some task, such as reading, the critical question is whether the artificial system performs the task in the same way as humans do. This requires the investigator to compare the nature and time course of the internal processes in humans to those of the artificial system. Only if the internal processes are similar can we say that the artificial system is a good model of the human system.

As students of reading processes, we are interested in whether a model holds up to its experimental tests and whether practical applications can be derived from the model. Although the model helps us understand the behavior, we also want it to help us modify it. An obvious example is whether a model proves useful in an assessment of reading skills and the development of reading instruction. A model's adequacy in providing assessment and guiding instruction

provides additional tests of the model. If a model assumes that a reader must have some knowledge of spelling constraints (orthographic structure) in order to recognize words for rapid reading, then readers without this knowledge should reveal reading problems. Acquiring this knowledge should lead to rapid reading if no other deficits exist.

The student of reading may not readily discover any consistent approaches to building and testing models in research on reading. The prescriptions of a theory of model building and testing are not always apparent in actual practice. Each substantive area of investigation has a unique history and state of the art; the scientist cannot easily avoid these influences. In this chapter, I discuss model building and testing in four or five substantive areas of reading research. Given the relatively advanced state of the art in letter and word recognition, this domain offers an ideal research environment for the discussion of model building and testing in reading research. My goal is to present to the student some tools for building and testing models in the context of actual investigation. The present framework for research is highly biased, as any informative and internally consistent framework must be. For alternative approaches to the study of reading, the student is referred to Guthrie and Hall and to Kamil in this volume.

Before initiating our discussion of specific models, it is necessary to mention the restricted domain of the models we will discuss. I am using the term *psychological model* to signify a formal description of some psychological phenomenon. In addition, the models to be discussed are concerned with information-processing characteristics of reading: how the reader travels from one state of knowledge to another by way of the printed medium. Unfortunately, there has been very little development of formal models of personality, social, cultural, and affective contributions to reading. Models within the genre of information processing may have *nothing* to say about what or why people read. Motivation, goals, and interests are critical to a complete understanding of reading, as they are in all other aspects of psychological functioning. Our modest, although not easily achieved, goal is to understand perceptual and cognitive functioning unique to reading. After some mastery of this domain, it will be necessary to extend the analysis to include the affective dimensions, accounting for why or what we read.

LETTER RECOGNITION

The role of model building in reading can be illuminated in the substantive area of letter recognition. How does a reader recognize some squiggles on the page as a familiar letter? There have been two basic approaches in building models of letter recognition: templates and features. Palmer (1978) provides an informative analysis of the properties of template and feature theories. He points out that the two theories are not easily distinguished. I believe the important difference between the two theories is their differences in explanatory power. In the template model, the implicit assumption is that letters are essentially Gestalt units that cannot be further analyzed or reduced in terms of other characteristics. Therefore, a sequence of letters in a text would consist of a sequence of indivisible patterns. Visual analysis of these patterns would be limited to some kind of scheme in which the reader would match a series of templates stored in memory with the visual pattern of each letter. A unique template is needed for every unique visual pattern. When a match is found, recognition occurs.

The central criticism of the template matching scheme is that it is very inflexible (Neisser, 1967). Size or orientation differences between the template and the visual pattern will baffle a direct matching between memory templates and visual information. It is possible to develop normalization routines, however, that allow modification of the visual pattern to put it into a format that corresponds more closely to the templates in memory. Given a set of normalization routines, variations in size and orientation are not necessarily an insurmountable obstruction for successful template matching.

A more critical problem with the template matching scheme is a metatheoretical[2] one. The template matching scheme provides little analysis or understanding of the phenomenon of interest. The template matching scheme may be interpreted as nothing more than a restatement of the problem. Although it postulates little theoretical machinery, the template matching scheme has no predictive power; for every item to be recognized, a template must be stored in memory. In order to predict the recognition of 26 letters, a template matching model must assume 26 templates. A good model must predict more than it assumes.

Each template must be considered a free parameter since the template matching model does not predict the nature of the templates. Although the concept of a free parameter is not easily articulated, an example of its use usually provides a sufficient explanation. Every quantitative description of some state of affairs has a set of parameters. As an example, Fechner's (1860/1966) psychophysical law was formulated and expressed as

$$S = K \log_{10} I,$$

where S measures the magnitude of the sensation, I the intensity of the stimulus, and K is a constant of proportionality. In this model, K and I serve as parameters. The stimulus intensity I is directly measurable in that Fechner specified exactly how it should be measured. The value of K cannot be measured directly and therefore is a free parameter. In other words, Fechner's law cannot predict the magnitude of sensation, given the stimulus intensity, until some K value is assumed. In testing the law, the obvious question is what K value should be assumed. The answer is, the K value that optimizes the predictions of the law. Any other value would produce an unfair test of the law because it would always be possible to find a K value that would violate the predicted relationship. More specifically, we want to find the K value that minimizes the differences between the observed and predicted values. For Fechner's law, a simple mathematical solution is possible; for more complicated models, iterative computer routines (e.g., Chandler, 1969) are available.

The predictive power of a model is determined by evaluating how much can be predicted relative to how much has to be assumed. To return to Fechner's law, it cannot be disproven if it is tested at only one stimulus intensity. Regardless of the sensation S, some value of K exists to predict it exactly. This would not

be the case with two or more levels of stimulus intensity. To the extent that a single free parameter K can predict the sensation at a large number of stimulus intensities, we gain confidence in the model. If Fechner's law holds for three or four intensity levels, we are somewhat impressed; if it holds across the complete range of audible intensities, we are very impressed. The student should be warned that there is no convenient measure of how well a model describes some set of outcomes relative to the number of free parameters needed to predict those outcomes. It is difficult to choose between two models when the model with the more accurate description also requires more free parameters.

For each unique letter to be recognized in the template matching model, a unique template is assumed, which requires another free parameter. Given this situation, the model has very little predictive value for the phenomenon of interest. One goal of a theory of letter recognition would be to devise a scheme for letter recognition that would allow the recognition of more letters than there are templates. In order to accomplish this, some theoretical process must be formalized to allow letter recognition on the basis of a relatively small amount of memory storage.

One scheme that accomplishes efficient recognition is called feature analysis and was primarily inspired by work in linguistics and artificial intelligence (Jakobson, Fant, & Halle, 1961; Selfridge, 1959). In linguistics Trubetzkoy and his colleagues of the Prague school questioned the idea that spoken language consists of minimal units called phonemes. Phoneme units could not be further divided, and therefore some kind of template matching scheme would be required for their recognition. Trubetzkoy and his colleagues, however, assumed that the phoneme units could be further analyzed in terms of distinctive features that represent similarities and differences with respect to other phonemes. The goal of this theoretical perspective was to analyze all the phonemes of a given language in terms of a small set of distinctive features. To the extent that the number of distinctive features was significantly less than the number of phonemes, the theory would have strong predictive value.

There are two important dimensions of a model of feature analysis of letter recognition. These are the structural aspects and process aspects of the model. Miller and Johnson-Laird (1976) propose that an important psychological problem is to characterize the *contents* of the perceptual world to account for other psychological phenomena. This problem would be addressed by the structural dimension of a model. Essentially, the world of experience is characterized in terms of variations in the stimulus world. In terms of a model of letter recognition, the goal would be to determine which characteristics of written letters are important in their recognition. The process component of a model is an attempt to account for the *generation* of the perceptual world, rather than the specific contents of the perceptual world. In this case, the emphasis would be, not on the characteristics of the letters that are critical for letter recognition, but on how the perceived characteristics are evaluated and integrated together in order to arrive at letter recognition. The goal would be to articulate rules for combining the perceived characteristics to produce the desired psychological outcome. Both the contents and the generation of the perceptual experience are important issues for a theory of letter recognition. A review of the visual features used in letter recognition and the processes involved is given by Massaro and Schmuller (1975). More recent work in this area can be found in Oden (1979), Naus and Shillman (1976), and Massaro (1979b).

WORD RECOGNITION

Disagreements in scientific endeavor prove productive if the researchers follow Platt's advice and test between opposing views. One question debated since the inception of reading-related research (Cattell, 1886; Gibson & Levin, 1975; Huey, 1908/1968; Massaro, 1975b) is to what extent word perception[3] is mediated by letter perception. According to one view, word perception follows naturally from perception of the letters in the word and the appropriate knowledge in the mind of the reader. The opposing view claims that a word is perceived as a unique configuration. These two theories have at various times served as the basis for the phonics and sight-word methods of reading instruction (Chall, 1967; Smith, 1971).

How does the scientist proceed to test between these views? First, each view must be explicitly developed in the form of a specific model. It is not possible to test between the two general statements given in the preceding paragraph, and therefore these statements must be transformed into specific models. Given a specific model, conclusions can be derived and can be checked for internal consistency and the possibility of being tested. The models derived from the different views should be contrasted to see if the conclusions are significantly different from one another. In many cases, two apparently discrepant views are transformed into very similar models when they are precisely formulated. If the two models make opposing conclusions, the scientist should still evaluate whether the conclusions are worth testing in terms of what knowledge would be gained. If the answer is positive, the researcher faces the challenge of mapping these conclusions into an experimental test.

I followed the above procedure in testing between the two theories of word perception (Massaro, 1979a). The theories were explicitly developed in a generally accepted view of word perception in reading. In this view, a printed pattern is first transduced by the visual receptor system, and featural analysis makes available a set of visual features. The word-perception process combines this visual information with other nonvisual knowledge and arrives at a perceptual experience of the printed pattern. The two theories seemed to differ in terms of whether or not the feature analysis of each letter is independent of the orthographic context imposed by the surrounding letters. If letters mediate word perception, feature analysis of the letters should not be modified by orthographic context. If a word is a unique configuration, feature analysis should be modified by orthographic context. A critical test between the two classes of models, therefore, could ask whether orthographic context influences feature analysis. The first theory says that orthographic context and featural analysis are independent; the second theory says that they are not. These will be referred to as independence and nonindependence theories.

Current models of word recognition were classified in terms of whether featural analysis and orthographic structure make independent contributions to letter perception. Consistent with the advice of Broadbent (1973), the goal of the experiment was to obtain results that would be probable under one class of models and improbable under the other class of models. The contrasting theories can be clarified by an analysis of recent formulations of specific models. Nonindependence views have taken the form of feature-analysis differences, higher-order units, and hypothesis-testing mechanisms.

A letter in a word is recognized more often than a letter presented alone or presented in a nonword under the same conditions (Johnston & McClelland, 1973; Reicher, 1969; Thompson & Massaro, 1973). This result obtains even when a simple guessing advantage for words is precluded by constraining the response alternatives. One of the primary interpretations of these findings is that the word context enhances the visual feature analysis of the component letters. If feature analysis were a limited-capacity process and subject to attentional control, then it might be expected that the familiarity of a word pattern would produce better feature analyses of the component letters. After an exhaustive and careful review of the literature, Krueger allows the possibility that the familiarity of orthographic context might influence the feature-analysis stage (Krueger, 1975, Table 1). According to this interpretation, readers would have a greater number of features and/or a better resolution of the features in a word than in a nonword context. This would mean that readers should be able to report more accurate detail about the visual characteristics of the letters in a word than in a nonword context.

In the second kind of nonindependence models, higher-order units intervene in the processing sequence and change the features and/or featural analyses employed. As an example, some models assume specific memory codes for spelling-pattern units, and some visual features are defined in terms of these higher-order units (Juola, Leavitt, & Choe, 1974; LaBerge & Samuels, 1974; Neisser, 1967; Smith & Haviland, 1972; Taylor, Miller, & Juola, 1977; Wheeler, 1970). As an example, the frequent spelling pattern *st* could function as a perceptual unit, and the top extent of the *s* and the horizontal cross of the *t* might function as a single supraletter feature. In this case, perception of the spelling pattern *st* might be easier than perception of either letter presented alone, since the presence of supraletter features gives the reader additional visual information (Wheeler, 1970).

In contrast to assuming a facilitative effect of higher-order units, other models in this set assume that word context can override the perception of the letters that make it up. In Johnson's (1975) model, for example, the feature set assigned to a letter sequence is determined by the entire sequence and higher-order memorial feature sets. In this case, a letter is assigned different features in different orthographic contexts. Johnson's model makes the prediction that the higher-order memorial feature sets camouflage the component letters and that therefore readers may know less about the visual characteristics of a particular letter in a word than that same letter presented alone or in a nonword string.

A final class of nonindependence models assumes that expectancy and hypothesis testing play an important role in the initial stage of visual analyses. In hypothesis-testing models, the current hypothesis directs feature analysis (Goodman, 1976a, 1976b; Wheeler, 1970). In the hierarchical feature-testing model of Wheeler (1970), for example, the detection of some features guides the detection of other features. Osgood and Hoosain (1974) have also made this kind of nonindependence assumption in their view of word perception. Their specific idea of nonindependence involves the derivation of the meaning of the word to feed back and influence the information passed on by peripheral sensory processes (feature detection in the present framework). This assumption is similar to an observing response model (Broadbent, 1967) in which the feature detection is biased by context and the solution given by Erdelyi (1974) for findings in the perceptual defense literature. The distinguishing attribute of this class of models is that feature processing is directly dependent on the guiding hypothesis; for example, a hypothesis that the word is *Philadelphia* might guide the feature analysis to test for an initial capital letter and a relatively long word. Accordingly, the visual features passed on by feature analysis vary as a function of the guiding hypothesis. The reader should know more about the visual characteristics of the segment of word that are relevant to the current hypothesis and less about the visual characteristics that are not relevant.

The current analysis makes transparent that nonindependence views of reading do not make the same predictions about how visual feature analysis is modified by orthographic context. However, in contrast to independence models, all nonindependence models assume that orthographic context changes the visual information passed on by feature analysis to the next stage of word recognition process.

Independence models follow in the tradition of Morton's (1969) logogen model in which higher-order context and visual information provide independent contributions to word recognition. Broadbent's (1971) model of the biasing effects of context and probability also qualifies as an independence model, since changes in these variables do not modify the intake of visual information. Although the issue of orthographic context was not addressed in these models, another independence model has been articulated in terms of word recognition as a direct consequence of featural information and orthographic context (Massaro, 1973, 1975a; Thompson & Massaro, 1973). The model was developed on the basis of experiments carried out using variants of the Reicher paradigm (Reicher, 1969; Wheeler, 1970). Subjects were presented with either a word or a single letter for a short duration, followed immediately by a masking stimulus and two response alternatives. The response alternatives would both spell words in the word condition; for example, given the test word *WORD*, the response alternatives *-D* and *-K* would be presented. The corresponding letter condition would be the test letter *D* followed by the response alternatives *D* and *K*. Performance was about 10 percent better in the word than in the single-letter condition.

Given the two-alternative forced-choice control, it was argued that the reader was able to utilize orthographic context to eliminate possible alternatives during the perception of the test display before the onset of the masking stimulus (Thompson & Massaro, 1973). As an example, given recognition of the context *WOR* and a curvilinear segment of the final letter, the reader could narrow down the alternatives for the final letter to *D*, *O*, and *Q*. Given that *O* and *Q* are orthographically illegal in the context *WOR-*, the letter *D* represents an unambiguous choice. The reader will therefore perceive the word *WORD* given just partial information about the final letter. If the reader has recognized the same curvilinear segment in the corresponding letter condition, however, any of the three letters (*D*, *O*, and *Q*) are still possible, and the perceptual synthesis will result in *D* only one out of three times. What is critical in this analysis is that a word advantage is obtained even though the visual information available to the perceptual process is equivalent in the word and letter conditions. The orthographic context of the word simply provides an additional but independent source of information. The featural information available to the recognition process does not change with changes in orthographic context. In this view, although orthographic context facilitates word perception, it does not modify the feature analysis of the printed pattern.

An experiment was carried out to test between the two general classes of nonindependence and independence models. The logic of the experiment to test between these two theories centered on the question whether context and featural processing make independent contributions to letter perception. The experiment involved the independent variation of the visual information about a letter and its orthographic context in a letter perception task. Consider the lowercase letters *c* and *e*. It is possible to gradually transform the *c* into an *e* by extending the horizontal bar. To the extent the bar is long, the letter resembles *e* and not *c*. If the letter is now presented as the first letter in the context *-oin*, the context would support *c* but not *e*. Only *c* is orthographically legal in this context because three consecutive vowels would violate English orthography. This condition is defined as *e* illegal and *c* legal ($\bar{e} \wedge c$). Only *e* is valid in the context *-dit* since the cluster *cd* is an invalid initial pattern in English. In this case, the context *-dit* favors *e* ($e \wedge \bar{c}$). The contexts *-tso* and *-ast* can be considered to favor neither *e* nor *c*. The first remains an illegal context whether *e* or *c* is present ($\bar{e} \wedge \bar{c}$), and the second is orthographically legal for both *e* and *c* ($e \wedge c$).

The experiment factorially combined six levels of visual information with these four levels of orthographic context, giving a total of 24 experimental conditions. The bar length of the letter took on six values going from a prototypical *c* to a prototypical *e*, as shown in Figure 5.1. In addition, the figure shows that the test letter was also presented at each of the four letter positions in each of the four contexts. A single test string was presented for a short duration, followed after some short interval by a masking stimulus composed of random letter features. In all cases, the subject indicated whether an *e* or *c* was presented in the test string.

Nonindependence theory assumes that featural processing is dependent on context. One direct measure of featural processing in the present experiment is the degree to which the reader can discriminate the bar length of the test letter. This discrimination can be indexed in the present experiment by the degree of differential responding to the successive levels of the bar length of the test letter. Better featural processing of the test letter is assumed to occur to the extent the subject responds *e* to one length and *c* to another. In the ($e \wedge c$) context, both letters spell words, whereas neither letter spells a word in the ($\bar{e} \wedge \bar{c}$) context. If the word context influences featural processing, as assumed by nonindependence theory, then the discrimination of bar length should differ in word and nonword contexts. If it does not, this would provide

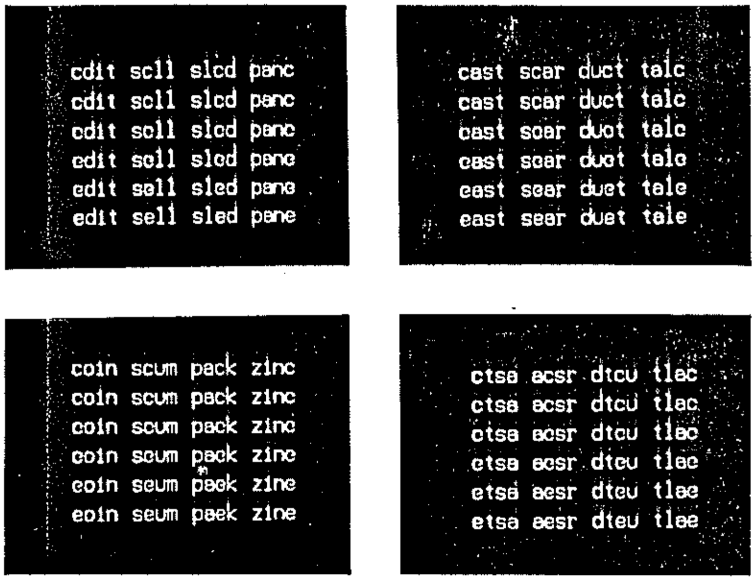

FIGURE 5.1 The 96 test items generated by the factorial combination of six bar lengths of the test letter, four serial positions of the test letter, and four orthographic contexts. (*After Massaro, 1979a.*)

a critical failure of nonindependence theory. The results indicated that context did not influence the discrimination of bar length, a critical failure of the nonindependence theory (Massaro, 1979a).

A qualitative test of the independence model was not available, and it was necessary to develop the model quantitatively. I now describe the quantitative model and how it was tested. In the experimental task, two independent sources of information are available: the visual information from the critical letter and the orthographic context. The first source of information can be represented by V_i where the subscript i indicates that V_i changes only with bar length. For the *e-c* identification, V_i specifies how much *e*-ness is given by the critical letter. This value lies between zero and one and is expected to increase as the length of the bar is increased. With these two letter alternatives it is reasonable to assume the visual information supporting c is simply one minus the amount of *e*-ness given by that same source. Therefore, if V_i specifies the amount of *e*-ness given by the test letter, then $(1 - V_i)$ specifies the amount of *c*-ness given by that same test letter.

The orthographic context provides independent evidence for e and c. The value C_j represents how much the context supports the letter e. The subscript j indicates that C_j changes only with changes in orthographic context. The value

of C_j lies between zero and one and should be large when e is legal and small when e is illegal. The degree to which the orthographic context supports the letter c is indexed by D_j and is independent of the value of C_j. The value of D_j also lies between zero and one and should be large when c is legal and small when c is illegal.

Faced with two independent sources of information, the reader evaluates the amount of *e*-ness and *c*-ness from these two sources. The amount of *e*-ness and *c*-ness for a given test display can therefore be represented by the conjunction of the two independent sources of information:

$$e\text{-ness} = (V_i \wedge C_j) \tag{1}$$

$$c\text{-ness} = ((1 - V_i) \wedge (D_j)). \tag{2}$$

The *e*-ness and *c*-ness values given by both sources can be determined once conjunction is defined. Research in other domains has shown that a multiplicative combination provides a much better description than an additive combination (Massaro & Cohen, 1976; Oden, 1979; Oden & Massaro, 1978). Applying the multiplicative combination, Equations (1) and (2) are represented as:

$$e\text{-ness} = V_i \times C_j \tag{3}$$

$$c\text{-ness} = (1 - V_i) \times (D_j). \tag{4}$$

A response is based on the *e*-ness and *c*-ness values: a choice of e is assumed to be made by evaluating the degree of *e*-ness relative to the sum of *e*-ness and *c*-ness values. This choice rule is a direct application of Luce's (1959) choice axiom. The probability of an e response, $P(e)$, is expressed as

$$P(e) = \frac{V_i C_j}{V_i C_j + (1 - V_i)(D_j)}. \tag{5}$$

To derive $P(e)$ for the four orthographic contexts, the central assumption about context is that a given alternative is supported to the degree x by a legal context and to the degree y by a illegal context, where $1 \geq x \geq y \geq 0$. The values x and y do not have subscripts since they depend only on the legality of the context. Therefore, C_j is equal to x when e is legal in a particular context and equal to y when e is illegal. Analogously, D_j is equal to x when c is legal in a particular context and equal to y when c is illegal.

Given the context with e legal and c illegal,

$$(e \wedge \bar{c})\text{: } P(e) = \frac{V_i x}{V_i x + (1 - V_i) y} \tag{6}$$

since the *e*-ness is given by $V_i x$ and the *c*-ness by $(1 - V_i)y$. Analogous expressions for the other three contexts are

$$(e \wedge c)\text{: } P(e) = \frac{V_i x}{V_i x + (1 - V_i) x} = V_i \tag{7}$$

$$(\bar{e} \wedge \bar{c}): P(e) = \frac{V_i x}{V_i x + (1 - V_i) y} \quad (8)$$

$$(\bar{e} \wedge c): P(e) = \frac{V_i y}{V_i y + (1 - V_i) x} \cdot \quad (9)$$

Equations (6) and (9) predict an effect of context to the extent a legal context gives more evidence for a particular test letter than does an illegal context (i.e., to the extent $x > y$). A second feature of this model is that $P(e)$ is entirely determined by the visual information when the context supports both or neither of the test alternatives; Equations (7) and (8) both predict $P(e) = V_i$.

This form of the independence model was tested against the observed results. Figure 5.2 gives the observed and predicted values. The results showed large effects of bar length and orthographic context on the identification of the

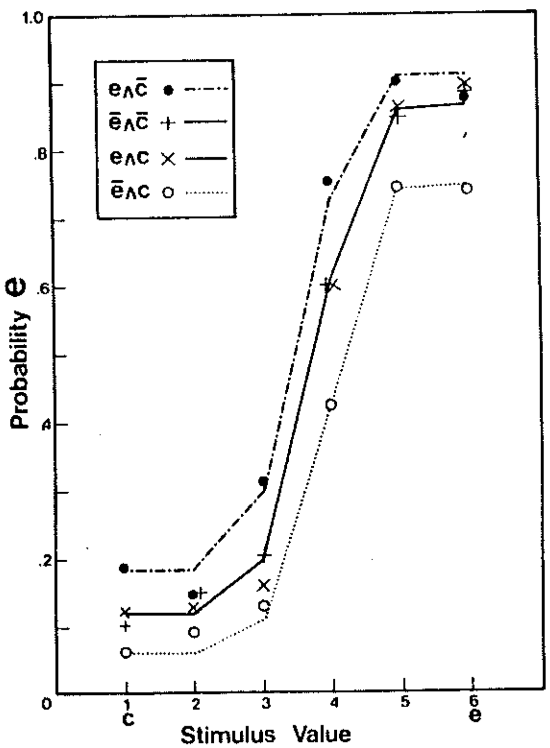

FIGURE 5.2 Predicted and observed probabilities of an *e* response as a function of orthographic context and stimulus value.

test letter. In order to fit the model to the data, it was necessary to estimate six values of V_i for each level of bar length of the critical letter and an x value for a legal context and a y value for an illegal context. The parameter values were estimated using the iterative routine STEPIT by minimizing the deviations between predicted and observed values (Chandler, 1969).

The model provided a good description of the results, considering that 24 independent observations are predicted with eight free parameters. In addition, the parameter estimates are psychologically meaningful. The value of V_i increased with increases in the length of the bar of the critical letter. The values were .11, .11, .19, .60, .85, and .86 for the six levels. The value of x was .76 for legal context, and the value of y was .40 for the illegal context.

Following the caveat of Hume, Popper, and Platt, the results do not verify the independence model; they simply corroborate it. Nevertheless, these results were sufficient to falsify the predictions of nonindependence theories. The proper conclusion is that there is no evidence to reject the idea of the independence of visual information and orthographic context in word perception in reading. Some theorists may not want to abandon the nonindependence view because of this single failure to find a difference where one should have occurred. The theorist may want to interpret the negative result as a sampling error; the null result could have been due to sampling variability overwhelming the underlying difference between the experimental conditions. This view is difficult to defend because there were large and systematic effects of the independent variables in the experiment. In addition, identical results were obtained with another stimulus set of items. The nonindependence theorist may also argue that the experimental test is irrelevant since its method and procedure placed it outside the domain of the theory. Some justification for this interpretation would be necessary if it is to be taken seriously. As with most experiments, however, certain deficiencies may be discovered and additional research will be necessary to resolve the issue. The ideal of a critical experiment may not be obtainable, and we may have to be content with the gradual accumulation of knowledge in scientific endeavors. Hence the need for highly energetic and persevering researchers.

TEMPORAL COURSE OF READING

A concept central to models of reading processes is time. Reading is an event over time, and any model of reading processes must include time as a dimension of empirical and theoretical analysis. It follows that the time taken by certain reading tasks becomes a crucial dependent measure for students of reading. In addition, it was discovered that the early stages of visual information processing in reading could be terminated at various times after their initiation. In this situation, performance accuracy could be used to index the processing activity up to its termination. The first paradigm is called a reaction-time task and the second is called backward recognition masking.

Reaction Time

The goal is to build models of performance in these tasks while making the critical assumption of a correspondence between some processes in the task and processes occurring in natural reading. Reaction-time research is carried out in a relatively intuitive theoretical framework. The framework was first articulated over a hundred years ago by Donders, a Dutch scientist best known for his work in ophthalmology. Donders (1868–1869/1969) assumed that a simple performance task involved a strict sequence of processes between the presentation of some stimulus event and the initiation of some response on the part of the observer. Donders developed a theoretical and methodological paradigm in order to measure the time of these intervening processes. Donders took issue with the then commonly accepted notion that responses to stimuli were infinitely short. He cited Helmholtz's work, which allowed the measurement of the time between the presentation of some stimulus on the skin and an involuntary reflex to that stimulus. Donders reasoned that it would be possible to devise two experimental situations in which the processing required would be very similar, except that one additional process would be required in one experimental situation relative to the other. The additional time required by the task with the additional process would be the time taken by the additional process. This was called the subtractive method.

One of the first questions to be addressed was the time taken to recognize a stimulus. Donders developed two tasks to study the time for recognition. In the first task, the subject was required to make a predetermined response whenever any visual stimulus was presented. The second task required the subject to make the same predetermined response only when a *particular* visual stimulus was presented. Donders reasoned that the subject would simply have to detect the presence of a visual stimulus before initiation of the response in the first task, whereas the stimulus would also have to be recognized as the appropriate alternative before initiation of the response in the second task. Since Donders believed that if all other aspects of the task were held constant, such as the response required, then the difference in reaction times between the two tasks should be a pure measure of the recognition time of the stimulus as the appropriate alternative. Donders's results were very promising in that he observed that recognition time was on the order of one-twentieth of a second. Therefore, the results of the experiment tended to support both the general framework of the subtractive method for measuring the time for mental events and provided specific information about the time taken for recognition.

Donders and others extended this subtractive method to a variety of experimental conditions and provided quantitative measures of such mental processes as recognition, discrimination, and response selection.

Donders's tasks did not go unchallenged. A number of investigators at the turn of the century criticized the central assumption made in the subtractive method. They argued that the assumption that one mental process can be added to a task without affecting the time to complete other mental processes was not valid. Changing the task to add a single stage probably affected the time it took for other stages. These critics did not provide any evidence supporting this criticism but simply relied on the introspective reports of their observers. Although the criticism was without empirical foundation, investigators lost interest in Donders's subtractive method as a tool for studying mental processes.

Other research could have also contributed to the downfall of the subtractive method. A century ago, James McKeen Cattell reported a series of experiments that created difficulty for the idea of completely successive mental processes.

Of particular interest to us are Cattell's studies of the processing stages in reading. The subjects in Cattell's (1886, 1888/1947) experiment included Cattell himself, George Stanley Hall, John Dewey, and Joseph Jastrow—a pretty impressive bunch of readers. Cattell modified Donders's recognition task to make it continuous. Letters were printed on a rotating cylinder that could be viewed through a slit. The size of the slit and the speed of rotation of the cylinder could be varied. The task of the subject was to name the letters, one by one, as they were exposed in the slit. The main independent variable was the size of the slit. The dependent variable was the rotation speed of the cylinder required by the subject to manage the task without error. If the window allowed only one letter at a time to be exposed, subjects required 360 milliseconds (msec) per letter in order to perform the naming task. On the other hand, by increasing the size of the window to expose more than one letter, the time needed per letter decreased from 360 to 225 msec with increases between one and four letters. Further increases beyond four letters did not speed up the task.

Cattell used these results to argue against the strict sequence of intervening processes assumed by Donders. If processing was purely sequential in that the naming of the second letter could not begin until the naming of the preceding letter was completed, then increasing the size of the window to expose more letters should have made no difference. The fact that increasing the size of the window did facilitate performance provides evidence for overlapping psychological processes. The processing of a second letter can begin while the first is still being processed.

Although Cattell did not analyze his result in the framework of a formal model, three interpretations can be formalized. First, Donders's sequential model may still be correct in terms of the processing of any single letter, but the processing of the second letter can begin before the processing of the preceding letter can be completed. That is, it could be the case that the subject, having recognized the first letter, and beginning to select the response to the first letter, can proceed to recognize the second letter, even though the response selection stage is not completed. Therefore, with respect to any given letter, Donders's sequential stage is valid but there is an overlapping of processes for the different letters.

A second explanation of Cattell's results is that the subject did not process the letters in a completely serial manner. Two or more letters could be recognized in parallel or simultaneously. Similarly, the response selection could occur in parallel for two or more letters. Both explanations do not necessarily invalidate Donders's sequential model but simply show that the unit of analysis at each stage is not necessarily the single letter. Although subjects can process two or more letters simultaneously, it does not invalidate Donders's assumption that response selection cannot begin until recognition is completed.

A third example rejects the notion that response selection does not begin until recognition is complete. Partial recognition of the first letter could allow some response selection to take place. As an example, subjects may recognize the first letter as one of five alternatives and pass this information on to the response selection stage. The response selection stage could then proceed to make these five letters available, waiting for additional information from the recognition stage.

Donders's subtractive method was revived a century later by Sternberg's modification and formal development of the additive-factor method (Sternberg, 1969a, 1969b). The additive-factor method is a powerful tool for studying models of reading processes. Other developments (e.g., McClelland, 1979) even permit a formal analysis of the reaction-time results when the intervening processes are not perfectly sequential but overlap in time. Research and model testing using the additive-factor method can be found in Massaro (1975a), Meyer, Schvaneveldt, and Ruddy (1975), and Theios (1975). The value of breaking down reaction times into times for separate psychological processes is illustrated in the discussion of building and testing models of phonological mediation in reading.

Backward Recognition Masking

Work in backward recognition masking began about the same time as Donders's work using the reaction-time paradigm. Baxt, in 1871, was interested in the time required to recognize letters. Baxt (cited in Sperling, 1963) did not have the distinguished subjects available to Cattell, so he used himself. He presented himself with a display containing a number of letters for a very short duration of about five msec. The display of letters was followed by a bright light flash. Baxt assumed that the bright light flash would terminate any further processing of the letters. The dependent measure in the task was the number of letters that could be seen as a function of the time between the presentation of the display and the presentation of the bright light. Baxt apparently could see one additional letter for every additional ten msec of time between the display and the bright flash. Therefore, he estimated recognition time for each letter as ten msec.

Sperling (1963) set out to replicate these results. Rather than use a bright light as a means of terminating processing of the letters, Sperling used a noise field, which was a series of random bits of letters. However, instead of presenting the display for a fixed duration and varying the blank interval between the display and the noise mask, the display remained on during the interval before the onset of the noise mask. Sperling replicated Baxt's results exactly. For every increase in presentation time of ten msec, Sperling's subjects were able to read out an additional letter. Maximum performance, of course, was only about four or five letters because of what would be expected from a short-term memory limitation. Sperling's results rest on the assumption that the noise mask terminated further processing of the test display. Given the short interval between the test display and the noise mask, however, it is possible that the two stimuli were visually integrated, and the subject was actually perceiving the letters after the noise mask was presented. More recent experiments have shown that this is probably the case (Eriksen & Eriksen, 1971). Therefore, Baxt's experiment and Sperling's experiment do not necessarily provide information about the time it takes to recognize letters.

Although these experiments are problematic, experimental situations can be designed so that backward masking can provide information about the temporal course of letter and word recognition. The paradigm also allows a more direct assessment of the dynamic utilization of orthographic context by the reader. A couple of studies have looked at the perceptual advantage of words over single letters in the Reicher (1969) task. On each trial, either a single letter or a complete

word is presented. On both kinds of trials, a pair of one-letter alternatives is presented, one of which had appeared in the original stimulus. The subject's task is to state which of the two letters had appeared. For example, on word trials the subject might be presented with the word *WORD* for a very brief time. When the task was to name the fourth letter of the word, the alternatives *D* and *K* would be presented above the former location of the fourth letter. The subject would have to choose from one of these two alternatives. On single-letter trials, the letter *D* might be presented, and the subject would have to choose between the alternatives *D* and *K*. Reicher (1969) and Wheeler (1970) found that performance was about 10 percent more accurate for words than for single letters.

Johnston and McClelland (1973) discovered that the word-letter difference was a function of the masking condition and, therefore, the processing time available. Three kinds of visual displays were tested: word, single letter, and a single letter embedded in a nonalphabetic symbol, #. In one experiment the test stimulus was followed immediately by a pattern mask. The test stimulus was presented at a high figure-ground contrast so that a relatively clear image was seen for a short period. As discussed earlier and in Massaro (1975b), the pattern mask interferes with any further processing of the test stimulus. Words were recognized 14 percent better than single letters and single letters embedded in # symbols. Performance did not differ for the two kinds of single-letter trials. In a second experiment the test stimulus was presented at a lower luminance and was followed by a plain white field of the same luminance as the test field. In this case, the white field would not interfere with the image of the test stimulus, and the subject would see a fuzzy image for a relatively long period. Single letters were now recognized as well as words, and these displays were recognized 8 percent better than the letters embedded in # symbols.

Johnston and McClelland (1973) proposed three alternative interpretations for the results. First, different systems could be responsible for processing letters and words. More specifically, it could be the case that letters in a word are protected from the detrimental effects of the pattern mask. With reference to our previous discussion of the use of analogy in model building, an analogue for this explanation came from the literature attempting to show that speech is not processed by the same mechanisms used to process nonspeech sounds (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). A second interpretation of the results is based on the process of lateral masking, which is a degradation of perception of a letter caused by the adjacent contours of surrounding letters. If the same neural mechanism were responsible for backward masking and lateral masking, then the interference might follow a law of diminishing returns. This means that less lateral masking would be observed in the pattern-mask condition than in the white-field condition. Lateral masking would be responsible for the elimination of the word advantage in the white-field condition. This explanation accounts only for the differences in the word advantage between the two masking conditions; it does not account for the word advantage itself. The third explanation is that letters in words are processed faster than letters presented alone or in # symbols. The word advantage occurs when processing time is limited with the pattern mask, but not when there is no limit on processing time in the white-field masking condition.

Johnston and McClelland (1973) evaluated the differences between a pattern mask and essentially a no-mask condition. The white field following the display does not mask or terminate perceptual processing. To evaluate the effect of pattern masking on the word advantage, it is important to assess performance continuously during the temporal course of letter recognition. Therefore it is necessary to vary systematically the interval between the test and masking stimuli in the recognition task. Following this logic, Massaro and Klitzke (1979) replicated and extended Johnston and McClelland's (1973) study in order to assess how the contribution of orthographic context varies with the pattern masking interval. Letter, letter embedded in $'s, word, and nonword test displays were presented in Reicher-Wheeler task in a backward recognition masking experiment. The masking stimulus followed the test display after a variable silent interval, and on some trials no mask was presented.

The results revealed highly systematic effects of the test displays and the masking interval. For all test conditions, performance improved systematically with increases in the interval between the test and mask displays. What is of interest is the rate of improvement with increases in masking interval for each of the test displays. There was no difference between the nonwords and the letters embedded in $'s, and these will be referred to simply as nonword displays. However, rate of improvement is a theoretical construct and must be defined explicitly with a particular model of performance. For their model, Massaro and Klitzke used a model that had proved successful in a variety of other domains (Massaro, 1970, 1975a).

In this model, the quantitative formulation of the time course of recognition describes the temporal course of letter perception by the simple equation

$$d' = \alpha(1 - e^{-\theta t}), \qquad (10)$$

where d' represents the perceptibility of the letter in z units, α represents maximal perceptibility with unlimited processing time, t is the processing time measured from the onset of the test stimulus to the onset of a masking stimulus, θ is the rate of processing, and e is the natural logarithm. Perceptibility is assumed to be a negatively accelerating growth function of the processing time available. The value α is dependent on the properties of the visual display and the acuity of the visual system. If the perceptibility of a letter is lowered by the lateral masking of a neighboring letter, this should result in a lower value of α. The value of θ reflects the rate of perceptual processing—how quickly primary recognition occurs. The value of θ can be expected to be dependent upon process variables such as selective attention to a particular letter (Lupker & Massaro, 1979) and the degree to which the reader utilizes orthographic context to recognize a test letter.

The primary support for this description of the temporal course of letter recognition comes from backward recognition masking experiments. In this task, a target stimulus is presented for a short duration and is followed after some blank interval by a masking stimulus superimposed at the same location as the target. The basic finding is that recognition of the target improves with increases in the blank interval before the onset of the mask. Before applying the model to the issue of letter and word recognition, some clarification of backward masking

is necessary. There seems to be a general misunderstanding of what backward masking does and how it is interpreted in serial-stage models of information processing. One common misinterpretation is that a backward mask somehow works retroactively; it eliminates all perceptual information about the visual display. Thus, unless subjects manage to encode the display into abstract representation, they will not be able to report anything about what was presented. In the present view of backward masking, however, the mask simply terminates any further perceptual resolution; the resolution that occurs before the mask is presented is continuously passed on to the next stage. This information is not eliminated by the mask, although the mask may also function to interfere with the information at this level (Kallman & Massaro, 1979; Massaro, 1975a, Chapter 24). The gradual improvement in performance with increases in the masking interval reflects the continuous perceptual resolution before the mask is presented.

Massaro and Klitzke (1979) used the model of processing described by Equation (10) to predict systematic changes in the contribution of orthographic context with processing time in the backward recognition masking task. Two relevant factors in single-letter and nonword perception are assumed to operate identically in word perception. They are (1) the processing time available between the onset of a test letter and the onset of a masking stimulus and (2) the lateral masking of the perceptibility of a letter by its neighboring letters. In the model, lateral masking and orthographic context are identified with different parameters in Equation (10), describing performance as a function of processing time t. Adjacent letters that degrade the perceptibility of a neighboring letter should influence the α value for that letter. Adjacent letters that reduce the uncertainty of a given letter should increase θ, the rate of processing the information in that letter.

It is important to consider how orthographic context functions to reduce the uncertainty of a given letter and modify the rate of processing of that letter. Consider the case in which the single letter *c* is presented. Perceptual resolution of the letter is assumed to be a temporally extended and continuous process. As an example, when $t = 100$ msec, the letter may be resolved sufficiently to reduce the alternatives to *c*, *e*, and *o*; 200 msec might be needed to completely resolve the letter *c*. In the word test condition, the letter *c* may be presented in the context *coin*. In this case, because of lateral masking, *c* may not be resolved completely but may be seen at only 90 percent clarity even with unlimited processing time. Hence the α value for *c* would be lower in the word than in the single-letter condition. But the word context should also enhance the rate of processing the information in the letter *c*. If the context *-oin* is resolved completely and the alternatives for the first letter are limited to *c*, *e*, and *o*, no further visual processing is necessary given this visual information and the orthographic regularity of the written language. The strings *eoin* and *ooin* are illegal in the context *-oin*. Therefore, *c* is the only valid alternative for the first letter. Given limited visual information, *c* can be recognized exactly in the word context. When *c* is presented as a single letter, complete visual information is necessary for correct recognition.

The operation of lateral masking and the utilization of orthographic context leads to an expected interaction between the letter and word conditions and the processing time available. When processing time is maximal, rate of processing is unimportant, and the letter condition should show an advantage because the perceptibility of a letter in a word is reduced by lateral masking. With intermediate processing times, the advantage of orthographic context in the word condition should enhance the rate of processing of the test letter and therefore offset the deficit of lateral masking.

This model makes certain predictions for the Massaro and Klitzke experiment. First, performance on nonwords should be equivalent to that on letters in dollar signs at all masking intervals. Both have the disadvantage of lateral masking, and both possess no orthographic context. Second, letters in words should be recognized at a faster rate than letters in nonwords. Third, with very long processing times, single letters should be better recognized than letters in words because of the lateral masking in word strings. Fourth, and most important, the model also predicts that the disadvantage of lateral masking in words relative to single letters (different α's) can be overridden at the shorter masking intervals by the faster rate of processing letters in words (different θ's) because of the utilization of orthographic context.

The results of critical interest are the masking functions for each of the test strings. Since the model predicts d' values, it was necessary to compute observed d' values for the letter, word, and nonword conditions as a function of stimulus onset asynchrony, the time between the onset of the test display and the onset of the masking stimulus. The nonword and the letter-in-$ conditions were also averaged before the d' values were computed, since there was no significant difference between these conditions. The d' values were computed from the average percentage correct values. In this analysis, the hit rates correspond to the percentage correct values and one minus these values are the false alarm rates.

Figure 5.3 plots the observed d' values as a function of orthographic context and the stimulus onset asynchrony. A monotonic masking function was observed for each context condition, and the functions seem to rise at different rates and to different asymptotes in the different conditions. Although the perceptibility of a letter in a word is greater than a letter presented alone at short processing times, the opposite is the case for long processing times. This interaction is exactly what is predicted by the present formulation. In terms of the model, α should be larger for single letters than for words or nonwords because of the reduced signal-to-noise ratio due to the lateral masking of adjacent letters in the word and nonword conditions. The rate parameter θ, however, should be larger for words than for single letters or nonwords because orthographic constraints allow the reader to arrive at a decision about which letters are present at a faster rate in the word condition than in the single-letter condition. The clarity of the features for a letter in a word is less for a letter presented alone, but fewer of these features are necessary to arrive at a decision in the word condition than in the single-letter condition.

In fitting the model to the results, one α was estimated for the single-letter condition and another for the word and nonword conditions. The word and nonwords should have the same value of α since lateral masking should be equivalent in these two cases. With respect to the rate of processing, θ, letters in words should be processed at a faster rate than letters in nonwords or a letter presented

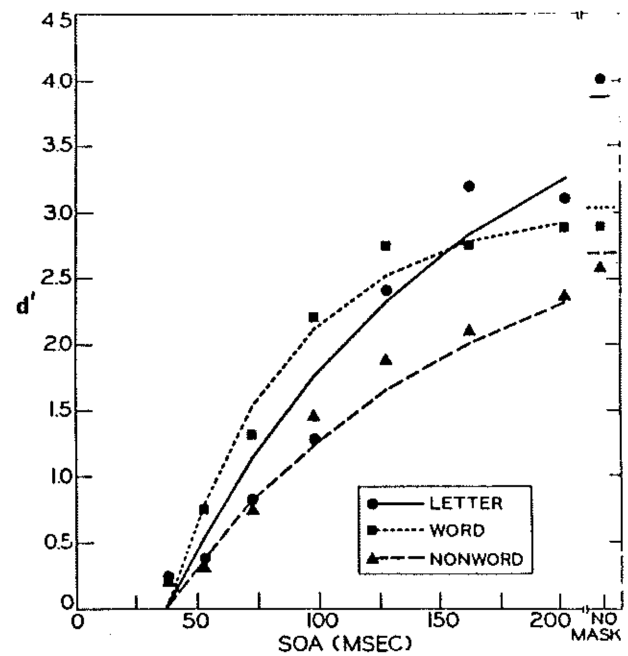

FIGURE 5.3 Predicted and observed d' values for the word, nonword, and single-letter conditions as a function of SOA.

alone. It follows that one value of θ should be estimated for words and another for letters and nonwords. Since the masking stimulus was more intense than the test stimulus, the mask would have a faster arrival time at the visual processing area than would the test stimulus. Therefore, it was also necessary to estimate a dead time, since the masking interval probably overestimated the true processing interval. Finally, it was necessary to estimate the duration that the display information was maintained in the visual processing center. This duration, t_D, gives the maximum processing time for very long masking intervals and the no-mask condition. If the masking interval exceeded t_D, then t_D was inserted in the equation.

The observed d' values were fit with the predictions of the model by estimating the six parameter values using the minimization subroutine STEPIT (Chandler, 1969). The α value for the letter alone condition was 4.32—significantly larger than the α value of 3.06 for words and nonwords. The θ value for words was 18.95—significantly larger than the θ value of 8.45 for letters and nonwords. The dead time was estimated to be 37 msec, and 266 msec was the estimated duration of preperceptual visual storage. The model provided a reasonably good description of the results, considering that 24 independent data points were described with just six parameter values. One test of the adequacy of the model is to see if the description would be improved when the constraints on α and θ were removed and two additional parameters are estimated. Therefore, if the α value for words differs from the α value for nonwords, and the θ value for letters differs from the θ value for nonwords, the description of the new model should be greatly improved. It was not, providing additional support for the original model.

The good description of the results by the model allows an assessment of the three interpretations offered by Johnston and McClelland (1973). First, it is not necessary to assume that letters and words were processed by qualitatively different mechanisms. The only difference for words is that readers have the added benefit of utilizing orthographic context. Second, it was not necessary to assume a tradeoff between backward masking and lateral masking mechanisms; the observed tradeoff is simply a natural outcome of the temporal course of processing. Third, words are processed faster than nonwords or letters alone due to the utilization of orthographic context. By providing results consistent with the third explanation, the experiment reveals that the first two interpretations are not necessary to explain the results of either the Johnston and McClelland (1973) or the Massaro and Klitzke (1979) experiments.

PHONOLOGICAL MEDIATION[4]

A very old question in reading-related research, one that is probably as old as reading itself, is whether the reader must translate print into some form of speech before meaning is accessed. This question can be formalized in terms of models in which speech mediation either does or does not occur in a derivation of meaning. Figure 5.4 presents a schematic diagram of both models. The top model assumes that phonological mediation must occur in order for the meaning of a message to be determined. In this model a letter string is presented, and the letters are identified by evaluating the visual information against feature lists of letters in long-term memory. The letters then are translated into sound or a soundlike medium by the spelling-to-sound correspondences of the language. One example of a spelling-to-sound rule would be that a medial vowel is usually pronounced as short unless it precedes a consonant followed by a final *e*. Thus, we have *fin* and *fine* or *fat* and *fate*. The speech code derived from spelling-to-sound rules is then used to access the lexicon in order to recognize the meaning of the word. The critical assumption of this model is that lexical access is achieved only by way of a speech code.

The bottom model in Figure 5.4 assumes that a speech code becomes available only after lexical access is achieved. The letters are identified in the same way as in the speech mediation model. However, the meaning of the letter string is determined by utilizing a visual code to achieve lexical access. The important assumption in this model is that the reader has information about what letter sequences represent what words. In this model the speech code becomes available only after lexical access is achieved. Given these two formal models, one would expect that it would be relatively easy to distinguish between

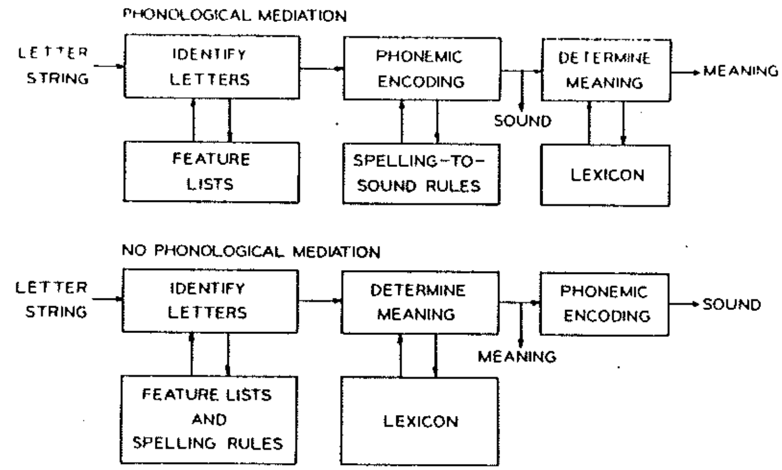

FIGURE 5.4 Two-stage models of the role of phonological mediation in reading.

them. Yet there is no consensus on which of these models is more likely to be correct.

Experiments have been done utilizing the time needed for lexical access in order to test between the models. Consider an experiment carried out by Gough and Cosky (1976). Subjects were asked to read aloud words that either obeyed or violated spelling-to-sound rules. One example of a word that violates a spelling-to-sound rule would be the word *give;* the vowel is not pronounced as long as it should be, as, for example, in the word *hive*. If subjects recognize words via speech mediation and utilize spelling-to-sound rules to achieve the speech code, then recognition of the word *give* should take longer than recognition of the word *hive*. Utilizing spelling-to-sound rules the reader would first interpret the letters *give* as [giyv] (rhymes with *hive*). Failing to achieve lexical access, a backup strategy would be initiated, and the short form of the vowel would be inserted giving the pronunciation [gIv]. In this case, lexical access would be achieved on the second try. Given the word *hive*, recognition would occur directly from the speech code [hiyv] produced by spelling-to-sound rules. The obvious hypothesis is that the pronunciation time for exception words, such as *give*, should be longer than the pronunciation times for regular words, such as *hive*. Gough and Cosky found that the pronunciation times for the exception words were in fact 27 msec longer than the pronunciation times for regular words. This result would seem to provide evidence for the speech mediation model.

Before these results can be interpreted as supporting the speech mediation model, however, it is necessary for the investigator to locate the differences in reaction time at the word-recognition stage of processing. The naming task requires a number of processing stages, and it is necessary to perform a stage analysis of the naming task. Naming a written pattern includes word-recognition and naming response operations. In terms of this analysis, reaction time between the onset of the written pattern and the onset of the spoken response is a composite of these two component times plus the times for other processes.

It is necessary to ask whether the reaction-time differences observed by Gough and Cosky are due to word recognition (lexical access), as they assume, or are due to naming response operations. It could be that lexical access time did not differ for exception and regular words but that the time for the subjects to program the naming response on achieving lexical access differed for exception and regular words. It could be that exception words are more difficult to pronounce once they are recognized and therefore require more time in the pronunciation task. Until the stage of processing is localized, these results cannot be taken as evidence for the speech mediation model. One possible control would have been to present the words auditorily and see if naming times differ in this situation. If they did not, this would provide some evidence that the differences in naming time were due to differences in time to achieve lexical access in the visual presentation condition.

Another way to test the differences between the exception and regular words, without confounding response selection and programming, would be to require a category judgment task. Here subjects could be asked to categorize the words, such as being nouns or verbs. The differences in the times to complete the categorization task would not be confounded with response processes because the response of categorization is identical for the exception and regular words. Following this logic, Bias (cited by McCusker, Hillinger, and Bias, 1981) found differences between exception and regular words using animal/nonanimal judgments. Further evidence against a response account of spelling-to-sound influences comes from a study in which both pronunciation and lexical decision[5] times were shorter for regular than for exception words.

Even if lexical access is slower for exception than for regular words, speech mediation may not be responsible. Differences between exception and regular words could be the result of differences in the times to process the letters of the words. There is good evidence that readers utilize orthographic structure to facilitate letter processing in word strings (Krueger & Shapiro, 1979; Massaro, 1979a). It could be that exception words have less orthographic structure than regular words, and letter recognition is therefore faster for regular than for exception words. Controlling for orthographic structure differences is difficult because an exact description of structure has not been validated. It is necessary to test for differences between these two classes of words in tasks that do not involve lexical access. As an example, a control experiment could have subjects search for a given target letter in regular and exception words. If the letter strings did not differ with respect to orthographic structure, then the time to search for a target letter in the strings would not differ for exception and regular words.[6]

SENTENTIAL CONTEXT

One of the oldest questions in reading-related research is the extent to which syntactic-semantic context influences word recognition. Many experiments have been carried out to demonstrate that sentential context influences word recognition. In contrast to the plethora of experimental work, there have been few models of how context works in word recognition. Gough, Alford, and Holley-Wilcox (1978) discuss a model that assumes the contribution of context is independent of the extraction of visual information from the word. This assumption is similar to the assumption of the independence of featural analysis and orthographic context in the independence model given in the section on word recognition. The Gough et al. model was formulated by Tulving, Mandler, and Baumal (1964). These authors proposed that the independence assumption made the following prediction: Assume that f is the probability of recognition of a word presented without any context, and c is the probability of recognizing the same word given only the context. If the visual information and context are processed independently, then p, the probability of recognizing the word, given both the stimulus presentation and context, should be

$$p = f + c - fc, \qquad (11)$$

which is the equation given by probability theory for describing the summation of two independent probability events.

Some of the original studies of the combination of sentential context and stimulus information in word recognition were performed by Tulving and his colleagues (Tulving & Gold, 1963; Tulving et al., 1964). Tulving et al. (1964) combined eight exposure durations with four context lengths in a word-recognition task in which a tachistoscopic presentation of a word followed the reading of the context. Subjects read either the last zero, two, four, or eight words of the context part of the sentence, and the test word was presented at either 20, 40, 60, 80, 100, 120, or 140 msec. This gives a total of 32 experimental conditions. Subjects were instructed to write down the test word and guess if they were not sure of their answer. They were told that the context words might be helpful in recognition of the test word.

Figure 5.5 presents the percentage of correct responses as a function of the duration of the test word; context length is the curve parameter. The context with two words is not presented because it produced results roughly identical to the context with four words. As expected, performance improved with increases in word duration. Performance also improved with increases in sentential context. The results show a larger contribution of sentential context when the exposure duration is intermediate and performance is neither very poor nor very good. This result indicates that context is most effective when subjects have some, but not relatively complete, featural information about the test word.

Tulving et al. (1964) tested the independence idea given by Equation (11) against the results of their experiment. A straightforward interpretation of Equation (11) would assume that the duration of the stimulus can influence f and that the number of context words can influence c. Therefore, eight parameter values of f and four parameter values of c must be estimated before the model can be evaluated. The estimates of f were assumed to be the probabilities of a correct response at the zero context condition. Similarly, the estimates of c were obtained from the response probabilities at the zero exposure duration. To test the model, the authors compared the observed probabilities under the 21 other experimental conditions to those predicted by Equation (11), given the appropriate estimates of f and c. The observed recognition probabilities were in every

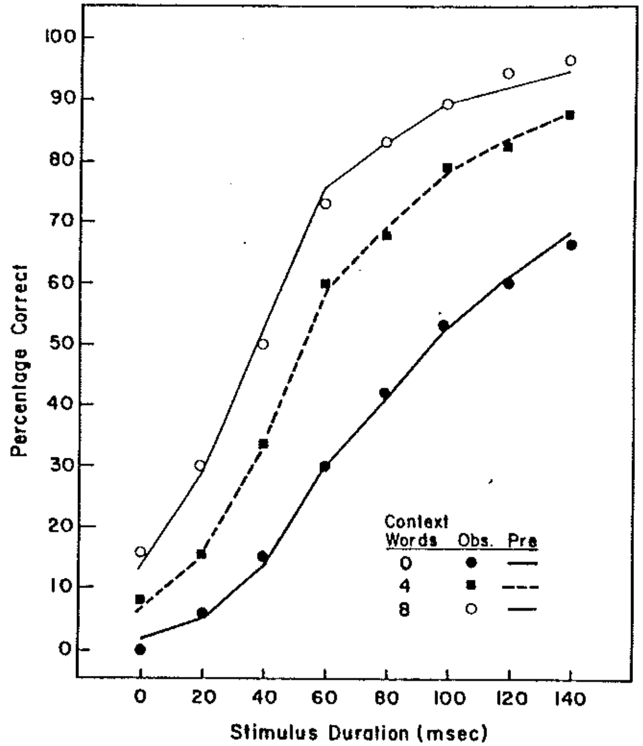

FIGURE 5.5 Observed (points) and predicted (lines) percentage correct identifications as a function of stimulus duration and the number of context words.

case larger than that predicted by the independence equation (11). Therefore, the idea of independence of stimulus information and context was rejected.

The rejection of independence may have been premature. As Gough et al. (1978) pointed out, the response probabilities at the zero context or zero duration conditions may not be valid estimates of c and f. All the observed data should be used to make the parameter estimates. Tulving et al. did not have a computer to estimate the parameters by minimizing the deviations between the predicted and observed recognition probabilities. A more critical problem in the application of the model is the implicit assumption that c is zero when no context is presented. It is possible that the experience of the readers in the experiment allowed them to utilize some knowledge about the set of target words even in the zero context condition. If this was the case, the model may also have failed for this reason.

It should also be stressed that the formulation of Tulving et al. is only one

instance of an independence model. A critical although implicit component of their formulation is that recognition via context and stimulus is all or none. Context either allows recognition of a word or it does not; it does not allow the contribution of partial information about the word. This is also true for stimulus information. This stands in sharp contrast to the idea of continuous information discussed in previous sections of this chapter. Moreover, the Tulving et al. formulation makes the implicit assumption that false information is not possible because there is no rule to apply if stimulus information triggers one alternative and context information another. In the fuzzy-logical model of Massaro (1979a), however, the reader has partial information from both context and stimulus presentation. The integration of these sources of partial information is analogous to the integration of letter information and orthographic context, discussed in the section on word recognition.

In the framework of the fuzzy-logical model, the reader has two sources of information about the test word; the visual information is indexed by V_i and the context by C_j. The reader evaluates both sources of information to arrive at the amount of support for a particular word alternative. The overall degree of support for the correct word can be indexed by g(correct word, S_{ij}), where S_{ij} is the stimulus condition corresponding to the ith level of visual information and the jth level of context. Combining the two sources of information would give

$$g(\text{correct word}, S_{ij}) = V_i \times C_j. \tag{12}$$

In Equation (12), V_i is the visual information, and C_j is the sentential context information. Before the overall degree of support for the correct alternative can be used to indicate the likelihood of a correct response, the support for all other alternatives must be taken into account. For simplicity, it seems reasonable to assume that there is a relative tradeoff between the correct and incorrect alternatives for each source of information. In this case, if V_i represents the degree of visual support for the correct alternative, $(1 - V_i)$ would represent the degree of support for all incorrect alternatives. Similarly, $(1 - C_j)$ would represent the degree of support for all incorrect alternatives given by sentential context. Therefore, the likelihood of a correct response should be equal to the support for the correct word relative to the total support for all alternatives:

$$P(\text{correct}) = \frac{V_i C_j}{V_i C_j + (1 - V_i)(1 - C_j)}. \tag{13}$$

In order to fit the model to the observed results, eight values of V_i and four values of C_j must be estimated as parameters in Equation (12). Figure 5.5 also presents the predictions of the fuzzy-logical model. The model provides a reasonably good description of the results, with a root mean square deviation of 2.1 percent.

We began with some general considerations in building and testing psychological models. These general considerations were an important aspect of the discussion of six or seven research problems within the domain of letter and word recognition in reading. Overall, the approaches seem to comply with the established frameworks for scientific endeavor. Scientists value theories that are parsimonious, that are capable of being falsified, and that make unique predictions in the domain of interest. We guard against confirmation biases by primarily having a highly interactive and competitive research area. No claims go unchallenged, and the theoretician is vulnerable to disconfirmation from other members in the field. Although one might believe that a good theory is one that is easily disproven, it is important to remember that rejection of a theory requires a viable alternative for significant progress in the field.

Our venture into model building and testing in reading research has revealed a parallel between reading and scientific endeavor. Both rely on basic perceptual, cognitive, and emotional processes of the individual. Reading, like scientific endeavor, begins with a large and unwieldy data base or knowledge structure. New information is processed and assimilated into the structure. Curiosities are developed, and hypotheses are generated, formulated, and tested against available data. Novel situations are explored and created in a continuous search for resolution and hope of completeness. To paraphrase Huey (1908/1968), to understand what we do in the scientific process would not only illuminate the process itself but might provide a fundamental understanding of those processes so characteristic of the only known species of readers.

NOTES

1. For our purposes, the terms *models* and *theories* can be used interchangeably, although models may be considered to be less general and more formal (precise) than theories.
2. The term *metatheory* is used to signify a philosophical framework that cannot be empirically tested. A metatheoretical issue is a point of logical debate; empirical evidence may come into play by evaluating the success of specific models derived within certain metatheories.
3. The terms *perception, recognition,* and *identification* are usually used interchangeably in reading research.
4. The terms *phonological* and *phonemic* are used to mean a speechlike code of a visible letter string.
5. A lexical decision involves classifying a letter string as a word or nonword without any overt pronunciation of the string.
6. One limitation with this control study is that lexical status also facilitates target search (Krueger, 1975; Massaro, Taylor, Venezky, Jastrzembski, & Lucas, 1980). If lexical access is achieved faster for regular than for exception words, we would expect faster target search for the regular than the exception words. An advantage of regular words in the target-search task would not be informative. Only if performance is equivalent in the target-search task for regular and exception words do we have evidence that these strings are roughly equivalent in orthographic structure. In addition, these results would indicate that lexical access was not achieved faster for the regular than for the exception words. It would follow that the differences between these two classes of words in the naming experiment were probably due to differences in a response rather than a recognition operation.

REFERENCES

Allport, D. H. On knowing the meaning of words we are unable to report: The effects of visual masking. In S. Dornic (Ed.), *Attention and performance VI.* Hillsdale, N.J.: Erlbaum, 1977.

Baron, J. The word-superiority effect: Perceptual learning from reading. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes.* Hillsdale, N.J.: Erlbaum, 1978.

Broadbent, D. E. Word-frequency effect and response bias. *Psychological Review,* 1967, *74,* 1–15.

Broadbent, D. E. *Decision and stress.* New York: Academic Press, 1971.

Broadbent, D. E. *In defense of empirical psychology.* London: Methuen, 1973.

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

Cattell, J. M. The time it takes to think. *Popular Science Monthly,* 1888, *32,* 488–491. (Reprinted in *James McKeen Cattell: Man of Science* (Vol. 2): *Addresses and formal papers.* Lancaster, Pa.: Science Press, 1947.)

Chall, J. *Learning to read: The great debate.* New York: McGraw-Hill, 1967.

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

Conant, J. B. *On understanding science.* New Haven: Yale University Press, 1947.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiano, M. D. *Pseudodiagnosticity (Acta Psychologica),* 1979, *43,* 111–121.

Donders, F. C. *On the speed of mental processes, 1868–1869.* Translated by W. G. Koster. In W. G. Koster (Ed.), *Attention and performance II (Acta Psychologica),* 1969, *30,* 412–431.

Erdelyi, M. H. A new look at the new look: Perceptual defense and vigilance. *Psychological Review,* 1974, *81,* 1–25.

Eriksen, C. W., & Eriksen, B. A. Visual perceptual processing rates and backward and forward masking. *Journal of Experimental Psychology,* 1971, *89,* 306–313.

Estes, W. K. Perceptual processing in letter recognition and reading. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 9). New York: Academic Press, 1978.

Estes, W. K. On the descriptive and explanatory functions of theories of memory. In L. Nilsson (Ed.), *Perspectives on memory research.* Hillsdale, N.J.: Erlbaum, 1979.

Fechner, G. T. *Elements of psychophysics* (Vol. 1), 1860. Translated and edited by H. E. Adler, D. H. Howes, & E. G. Boring. New York: Holt, Rinehart, & Winston, 1966.

Gibson, E. J., & Levin, H. *The psychology of reading.* Cambridge, Mass.: MIT Press, 1975.

Goodman, K. S. Behind the eye: What happens in reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading.* Newark, Del.: International Reading Association, 1976. (a)

Goodman, K. S. Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading.* Newark, Del.: International Reading Association, 1976. (b)

Gough, P. B., Alford, J. A., Jr., & Holley-Wilcox, P. *Words and contexts.* Paper presented at the annual meeting of the National Reading Conference, St. Petersburg Beach, Fla., December, 1978.

Gough, P. B., & Cosky, M. J. One second of reading again. In N. J. Castellan, Jr., D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive Theory* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1976.

Huey, E. B. *The psychology and pedagogy of reading.* New York: Macmillan, 1908. (Republished by MIT Press, 1968.)

Jakobson, R., Fant, C. G. M., & Halle, M. *Preliminaries to speech analysis: The distinctive features and their correlates.* Cambridge, Mass.: MIT Press, 1961.

Johnson, N. F. On the function of letters in word identification: Some data and a preliminary model. *Journal of Verbal Learning and Verbal Behavior,* 1975, *14,* 17–29.

Johnston, J. C., & McClelland, J. P. Visual factors in word perception. *Perception and Psychophysics,* 1973, *14,* 365–370.

Juola, J. F., Leavitt, D. D., & Choe, C. S. Letter identification in word, nonword, and single letter displays. *Bulletin of the Psychonomic Society*, 1974, *4*, 278–280.

Kallman, H. S., & Massaro, D. W. Similarity effects in backword recognition masking. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 110–128.

Krueger, L. E. Familiarity effects in visual information processing. *Psychological Bulletin*, 1975, *82*, 949–974.

Krueger, L. E., & Shapiro, R. G. Visual search for intact and mutilated letters through rapid sequences of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 657–673.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293–323.

LaBerge, D., & Samuels, S. J. (Eds.). *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, *74*, 431–461.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Lupker, S. J., & Massaro, D. W. Selective perception without confounding contributions of decision and memory. *Perception and Psychophysics*, 1979, *25*, 60–69.

McCain, G., & Segal, E. M. *The game of science*. Monterey, Calif.: Brooks/Cole, 1973.

McClelland, J. L. On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 1979, *86*, 287–330.

McConkie, G. W., & Rayner, K. The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics*, 1975, *17*, 578–586.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. Phonological recoding and reading. *Psychological Bulletin*, 1981, *89*, 217–245.

Massaro, D. W. Perceptual processing and forgetting in memory tasks. *Psychological Review*, 1970, *77*, 557–567.

Massaro, D. W. Perception of letters, words, and nonwords. *Journal of Experimental Psychology*, 1973, *100*, 349–353.

Massaro, D. W. *Experimental psychology and information processing*. Chicago: Rand McNally, 1975. (a)

Massaro, D. W. (Ed.). *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975. (b)

Massaro, D. W. A stage model of reading and listening. *Visible Language*, 1978, *12*, 3–26.

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

Massaro, D. W. Reading and listening. In P. A. Kolers, M. Wrolstad, & H. Bouma (Eds.), *Processing of Visible Language* (Vol. 1). New York: Plenum, 1979, 331–354. (b)

Massaro, D. W., & Cohen, M. M. The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *Journal of the Acoustical Society of America*, 1976, *60*, 704–717.

Massaro, D. W., & Klitzke, D. The role of lateral masking and orthographic structure in letter and word perception. *Acta Psychologica*, 1979, *43*, 413–426.

Massaro, D. W., & Schmuller, J. Visual features, preperceptual storage, and processing time in reading. In D. W. Massaro (Ed.), *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975.

Massaro, D. W., Taylor, G. A., Venezky, R. L., Jastrzembski, J. E., & Lucas, P. A. *Letter and word perception: Orthographic structure and visual processing in reading*. Amsterdam: North Holland, 1980.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word-recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V*. New York: Academic Press, 1975.

Miller, G. A., & Johnson-Laird, P. N. *Language and Perception*. Cambridge, Mass.: Belknap, 1976.

Mitroff, I. I. *The subjective side of science*. Amsterdam: Elsevier, 1974.

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, *76*, 165–178.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, 1977, *29*, 85–95.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology*, 1978, *30*, 395–406.

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

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.

Oden, G. C. A fuzzy logical model of letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *6*, 336–352.

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

Osgood, C. E., & Hoosain, R. Salience of the word as a unit in the perception of language. *Perception and Psychophysics*, 1974, *15*, 168–192.

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

Popper, K. *The logic of scientific discovery*. New York: Basic Books, 1959.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, *81*, 275–280.

Roediger, H. L., III. Memory metaphors in cognitive psychology. *Memory and Cognition*, 1980, *8*, 231–246.

Rumelhart, D. E. Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, N.J.: Erlbaum, 1977.

Selfridge, O. G. Pandemonium: A paradigm for learning. In *Symposium on the mechanization of thought processes*. London: Her Majesty's Stationery Office, 1959.

Smith, E. E., & Haviland, S. E. Why words are perceived more accurately than nonwords: Inference versus unitization. *Journal of Experimental Psychology*, 1972, *92*, 59–64.

Smith, F. *Understanding reading*. New York: Holt, Rinehart, & Winston, 1971.

Sperling, G. A model for visual memory tasks. *Human Factors*, 1963, *5*, 19–31.

Sternberg, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, *57*, 421–457. (a)

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 1969, *30*, 276–315. (b)

Taylor, G. A., Miller, T. J., & Juola, J. F. Isolating visual units in the perception of words and nonwords. *Perception and Psychophysics*, 1977, *21*, 377–386.

Theios, J. The components of response latency in human information processing. In S. Dornic & P. M. A. Rabbitt (Eds.), *Attention and Performance V*. New York: Academic Press, 1975.

Thompson, M. C., & Massaro, D. W. The role of visual information and redundancy in reading. *Journal of Experimental Psychology*, 1973, *98*, 49–54.

Tulving, E., & Gold, C. Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, 1963, *66*, 319–327.

Tulving, E., Mandler, G., & Baumal, R. Interactions of two sources of information in tachistoscopic word recognition. *Canadian Journal of Psychology*, 1964, *18*, 62–71.

Wheeler, D. D. Processes in word recognition. *Cognitive Psychology*, 1970, *1*, 59–85.

Woodworth, R. S. *Experimental psychology*. New York: Holt, Rinehart & Winston, 1938.

6 ASSESSMENT IN READING

Peter H. Johnston

This chapter examines the development and current status of the assessment of reading. While reading assessment is "as old as the first mother or teacher who questioned and observed a child reading" (Farr & Tone, in press), its documented development goes back only a short distance into the last century. The issues involved in the assessment of reading include (at least) the diverse areas of sociology, education, politics, philosophy, and all branches of psychology. The organization of such a chapter cannot be discrete, thus certain issues recur in different contexts in the various sections. The organization was selected on the basis of personal bias.

The chapter begins with a brief historical overview in which it is claimed that certain early developments set up a paradigm that all but determined our current assessment practices. As Haney (1981) comments, ". . . standardized tests appear to be social artifacts as much as scientific instruments" (p. 1030). The second section briefly addresses some often overlooked aspects of the construction and use of questions in assessment instruments. The third section describes some issues involved in the conflict between two opposing models of testing, and the fourth section discusses some issues involved in the concept of validity. The fifth section comments on trends in the deployment of assessment efforts.

Inertia and the various factors noted by Kuhn (1962) tend to allow us to maintain our basic paradigm without seriously evaluating it with a view to radical renewal. One constructive way for us to examine our current status is to challenge the assumptions of the existing paradigm, thus forcing a complete justification by those supporting the status quo. Consequently, the sixth section, in order to uncover some of the weaknesses in the state of the art, investigates where we might be now had we not pursued the current assessment approach. Finally, a brief summary is presented.

I am indebted to my colleagues Dick Allington, Peter Mosenthal, Fred Ohnmacht, David Pearson, and Jaap Tuinman for helpful comments on an earlier version of this paper, and particularly to David Pearson for getting me involved in it in the first place.