THE SCIENCE OF WORD RECOGNITION

THE SCIENCE
or how I learned to stop

text Kevin Larson
Advanced Reading Technology, Microsoft Corporation
translation Pavel Zelenka

Introduction

Evidence from the last 20 years of work in cognitive psychology indicate that we use the letters within a word to recognize a word. Many typographers and other text enthusiasts I've met insist that words are recognized by the outline made around the word shape. Some have used the term bouma as a synonym for word shape, though I was unfamiliar with the term. The term bouma appears in Paul Saenger's 1997 book Space Between Words: The Origins of Silent Reading. There I learned to my chagrin that we recognize words from their word shape and that "Modern psychologists call this image the 'Bouma shape.'"

This paper is written from the perspective of a reading psychologist. The data from dozens of experiments all come from peer reviewed journals where the experiments are well specified so that anyone could reproduce the experiment and expect to achieve the same result. This paper was originally presented as a talk at the ATypI conference in Vancouver in September, 2003.

The goal of this paper is to review the history of why psychologists moved from a word shape model of word recognition to a letter recognition model, and to help others to come to the same conclusion. This paper will cover many topics in relatively few pages. Along the way I will present experiments and models that I couldn't hope to cover completely without boring the reader. If you want more details on an experiment, all of the references are at the end of the paper as well as suggested readings for those interested in more information on some topics. Most papers are widely available at academic libraries.

I will start by describing three major categories of word recognition models: the word shape model, and serial and parallel models of letter recognition. I will present representative data that was used as evidence to support each model. After all the evidence has been presented, I will evaluate the models in terms of their ability to support the data. And finally I will describe some recent developments in word recognition and a more detailed model that is currently popular among psychologists.

Model #1: Word Shape

The word recognition model that says words are recognized as complete units is the oldest model in the psychological literature, and is likely much older than the psychological literature. The general idea is that we see words as a complete patterns rather than the sum of letter parts. Some claim that the information used to recognize a word is the pattern of ascending, descending, and neutral characters. Another formulation is to use the envelope created by the outline of the word. The word patterns are recognizable to us as an image because we have seen each of the patterns many times before. James Cattell (1886) was the first psychologist to propose this as a model of word recognition. Cattell is recognized as an influential founder of the field of psycholinguistics, which includes the scientific study of reading.

↑ **FIGURE 1** Word shape recognition using the pattern of ascending, descending, and neutral characters
↓ **FIGURE 2** Word shape recognition using the envelope around the word

Cattell supported the word shape model because it provided the best explanation of the available experimental evidence. Cattell had discovered a fascinating effect that today we call the Word Superiority Effect. He presented letter and word stimuli to subjects for a very brief period of time (5–10ms), and found that subjects were more accurate at recognizing the words than the letters. He concluded that subjects were more accurate at recognizing words in a short period of time because whole words are the units that we recognize.

Cattell's study was sloppy by modern standards, but the same effect was replicated in 1969 by Reicher. He presented strings of letters – half the time real words, half the time not – for brief periods. The subjects were asked if one of two letters were contained in the string, for example D or K. Reicher found that subjects were more accurate at recognizing D when it was in the context of WORD than when in the context of ORWD. This supports the word shape model because the word allows the subject to quickly recognize the familiar shape. Once the shape has been recognized, then the subject can deduce the presence of the correct letter long after the stimulus presentation.

The second key piece of experimental data to support the word shape model is that lowercase text is read faster than uppercase text. Woodworth (1938) was the first to report this finding in his influential textbook Experimental Psychology. This finding has been confirmed more recently by Smith (1969) and Fisher (1975). Participants were asked to read comparable passages of text, half completely in uppercase text and half presented in standard lowercase text. In each study, participants read reliably faster with the lowercase text by a 5–10% speed difference. This supports the word shape model because lowercase text enables unique patterns of ascending, descending, and neutral characters. When text is presented in all uppercase, all letters have the same text size and thus are more difficult and slower to read.

The patterns of errors that are missed while proofreading text provide the third key piece of experimental evidence to support the word shape model. Subjects were asked to carefully read passages of text for comprehension and at the same time mark any misspelling they found in the passage. The passage had been carefully designed to have an equal number of two kinds of misspellings: misspellings that are consistent with word shape, and misspellings that are inconsistent with word shape. A misspelling that is consistent with word shape is one that contains the same patterns of ascenders, descenders, and neutral characters, while a misspelling that is inconsistent with word shape changes the pattern of ascenders, descenders, and neutral characters. If test is the correctly spelled word, tesf would be an example of a misspelling consistent with word shape and tesc would be an example of a misspelling inconsistent with word shape. The word shape model would

predict that consistent word shapes would be caught less often than an inconsistent word shape because words are more confusable if they have the same shape. Haber & Schindler (1981) and Monk & Hulme (1983) found that misspellings consistent with word shape were twice as likely to be missed as misspellings inconsistent with word shape.

target word: **test**	error rate
consistent word shape **(tesf)**	13%
inconsistent word shape **(tesc)**	7%

↑ **FIGURE 3** Misspellings that are consistent with word shape are mi ssed more often

The fourth piece of evidence supporting the word shape model is that it is difficult to read text in alternating case. AlTeRnAtInG case is where the letters of a word change from uppercase to lowercase multiple times within a word. The word shape model predicts that this is difficult because it gives a pattern of ascending, descending, and neutral characters that is different than exists in a word in its natural all lowercase form. Alternating case has been shown to be more difficult than either lowercase or uppercase text in a variety of studies. Smith (1969) showed that it slowed the reading speed of a passage of text, Mason (1978) showed that the time to name a word was slowed, Pollatsek, Well, & Schindler (1975) showed that same-difference matching was hindered, and Meyer & Gutschera (1975) showed that category decision times were decreased.

Model #2: Serial Letter Recognition

The shortest lived model of word recognition is that words are read letter-by-letter serially from left to right. Gough (1972) proposed this model because it was easy to understand, and far more testable than the word shape model of reading. In essence, recognizing a word in the mental lexicon was analogous to looking up a word in a dictionary. You start off by finding the first letter, than the second, and so on until you recognize the word.
This model is consistent with Sperling's (1963) finding that letters can be recognized

at a rate of 10–20ms per letter. Sperling showed participants strings of random letters for brief periods of time, asking if a particular letter was contained in the string. He found that if participants were given 10ms per letter, they could successfully complete the task. For example, if the target letter was in the fourth position and the string was presented for 30ms, the participant couldn't complete the task successfully, but if string was presented for 40ms, they could complete the task successfully. Gough noted that a rate of 10ms per letter would be consistent with a typical reading rate of 300 wpm.

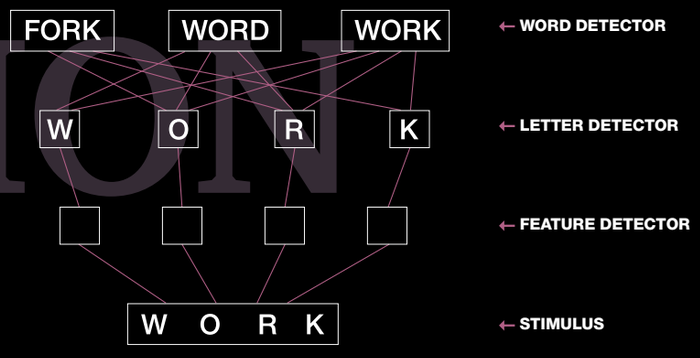
The serial letter recognition model is also able to successfully predict that shorter words are recognized faster than longer words. It is a very robust finding that word recognition takes more time with longer words. It takes more time to recognize a 5-letter word than a 4-letter word, and 6-letter words take more time to recognize than 5-letter words. The serial letter recognition model predicts that this should happen, while a word shape model does not make this prediction. In fact, the word shape model should expect longer words with more unique patterns to be easier to recognize than shorter words.

The serial letter recognition model fails because it cannot explain the Word Superiority Effect. The Word Superiority Effect showed that readers are better able to identify letters in the context of a word than in isolation, while the serial letter recognition model would expect that a letter in the third position in a word should take three times as long to recognize as a letter in isolation.

Model #3: Parallel Letter Recognition

The model that most psychologists currently accept as most accurate is the parallel letter recognition model. This model says that the letters within a word are recognized simultaneously, and the letter information is used to recognize the words. This is a very active area of research and there are many specific models that fit into this general category. I will only discuss one popular formulation of this model.

Figure 4 shows a generic activation based parallel letter recognition model. In this example, the reader is seeing the word work. Each of the stimulus letters are processed simultaneously. The first step of processing is recognizing the features of the individual letters, such as horizontal lines, diagonal lines, and curves. The details of this level are not critical for our purposes. These features are then sent to the letter detector level, where each of the letters in the stimulus word are recognized simultaneously. The letter level then sends activation to the word detector level. The W in the first letter detector position sends activation to all the words that have a W in the first position (WORD and WORK). The O in the second letter detector position sends activation to all the words that have an O in the second position (FORK, WORD, and WORK). While FORK and WORD have activation from three of the four letters, WORK has the most activation because it has all four letters activated, and is thus the recognized word.



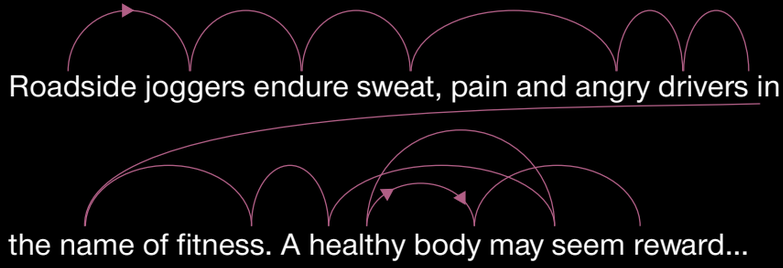
↑ **FIGURE 4** Parallel Letter Recognition

Much of the evidence for the parallel letter recognition model comes from the eye movement literature. A great deal has been learned about how we read with the advent of fast eye trackers and computers. We now have the ability to make changes to text in real time while people read, which has provided insights into reading processes that weren't previously possible.

It has been known for over 100 years that when we read, our eyes don't move smoothly across the page, but rather make discrete jumps from word to word. We fixate on a word for a period of time, roughly 200–250ms, then make a ballistic movement to another word. These movements are called saccades and usually take 20–35ms. Most saccades are forward movements from 7 to 9 letters,* but 10–15% of all saccades are regressive or backwards movements. Most readers are completely unaware of the frequency of regressive saccades while reading. The location of the fixation is not random. Fixations never occur between words, and usually occur just to the left of the middle of a word. Not all words are fixated; short words and particularly function words are frequently skipped. Figure 5 shows a diagram of the fixation points of a typical reader.

During a single fixation, there is a limit to the amount of information that can be recognized. The fovea, which is the clear center point of our vision, can only see three to four letters to the left and right of fixation at normal reading distances. Visual acuity decreases quickly in the parafovea, which extends out as far as 15 to 20 letters to the left and right of the fixation point.

Eye movement studies that I will discuss shortly indicate that there are three zones of visual identification. Readers collect information from all three zones during the span of a fixation. Closest to the fixation point is where word recognition takes place. This zone is usually large enough to capture the word being fixated, and often includes smaller function words directly to the right of the fixated word. The next zone extends a few letters past the word recognition zone, and readers gather preliminary information about the next letters in this zone. The final zone extends out to 15 letters past the fixation point. Information gathered out this far is used to identify the length of upcoming words and to identify the best location for the next fixation point. For example, in Figure 5, the first fixation point is on the s in Roadside. The reader is able to recognize the word Roadside, beginning letter



↑ **FIGURE 5** Saccadic eye movements

* Average saccade length and fixation times vary by language. The data presented here are for American English readers. While the values vary by language, it is remarkable that reading cognitive processes change so little from language to language.

information from the first few letters in joggers, as well as complete word length information about the word joggers. A more interesting fixation in Figure 5 is the word sweat. In this fixation both the words sweat and pain are short enough to be fully recognized, while beginning letter information is gathered for and. Because and is a high frequency function word, this is enough information to skip this word as well. Word length information is gathered all the way out to angry, which becomes the location of the next fixation. There are two experimental methodologies that have been critical for understanding the fixation span: the moving window paradigm and the boundary study paradigm. These methodologies make it possible to study readers while they are engaged in ordinary reading. Both rely on fast eye trackers and computers to perform clever text manipulations while a reader is making a saccade. While making a saccade, the reader is functionally blind. The reader will not perceive that text has changed if the change is completed before the saccade has finished.

Moving Window Study

In the moving window technique we restrict the amount of text that is visible to a certain number of letters around the fixation point, and replace all of the other letters on a page with the letter x. The readers task is simply to read the page of text. Interestingly it is also possible to do the reverse and just replace the letters at the fixation point with the letter x, but this is very frustrating to the reader. If just the three letters to the left and right of the fixation point are replaced with x, then reading rate drops to 11 words per minute. McConkie & Rayner (1975) examined how many letters around the fixation point are needed to provide a normal reading experience. Figure 6 shows a snapshot of what a reader would see if they are reading a passage and fixated on the second e in experiment. If the reader is provided three letters past the fixation point, then they won't see the entire word for experiment, and their average reading rate will be a slow 207 words per minute. If the reader is given 9 letters past the fixation point, they will see the entire word experiment, and part of the word was. With 9 letters, reading rate is moderately slowed. If the reader is given 15 letters past the fixation point, reading speed is just as fast as if there was no moving window present. Up to 15 letters there is a linear relation between the number of letters that are available to the reader and the speed of reading.

window size	sentence	reading rate
3 letters	An experimxxx xxx xxxxxxxxx xx	207 wpm
9 letters	An experiment wax xxxxxxxxx xx	308 wpm
15 letters	An experiment was condxxxxx xx	340 wpm

↑ **FIGURE 6** Linear relationship between letters available in moving window and reading rate.

window size	sentence	reading rate
3 letters	An experimxxx xxx xxxxxxxxx xx	207 wpm
1 word (3.7 lettres)	An experiment xxx xxxxxxxxx xx	212 wpm
9 letters	An experiment wax xxxxxxxxx xx	308 wpm
2 words (9.6 lettres)	An experiment was xxxxxxxxx xx	309 wpm
15 letters	An experiment was condxxxxx xx	340 wpm
3 words (15.0 lettres)	An experiment was conducted xx	339 wpm

↑ **FIGURE 7** Full word information does not improve reading rate.

From this study we learned that our perceptual span is roughly 15 letters. This is interesting as the average saccade length is 7–9 letters, or roughly half our perceptual span. This indicates that while readers are recognizing words closer to the fovea, we are using additional information further out to guide our reading. It should be noted that we're only using information to the right of our fixation point, and that we don't use any letters to the left of the word that is currently being fixated. In figure 6, where the user's fixation point is on the second e in experiment, if the word An is removed, it will not further slow reading rate. The moving window study demonstrates the importance of letters in reading, but is not airtight. The word shape model of reading would also expect that reading speed would decrease as word shape information disappears. The word shape model would make the additional prediction that reading would be significantly improved if information on the whole word shape were always retained. This turns out to be false.

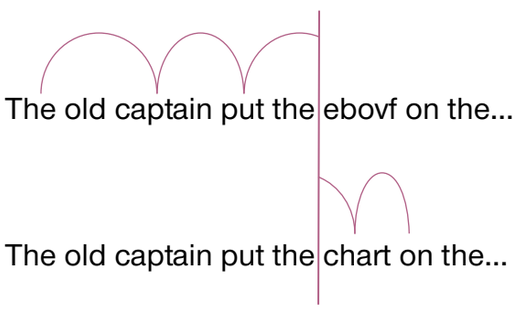
Figure 7 shows the reading rate when three letters are available. It is roughly equivalent to the reading rate when the fixated word is entirely there. That's true even though the entire word has 0.7 more letters available on average. When the fixated word and the following word are entirely available, reading rate is equivalent to when 9 letters are available. Reading rate is also equivalent when three words or 15 letters are available. This means that reading is not necessarily faster when entire subsequent words are available; similar reading speeds can be found when only a few letters are available.

Pollatsek & Rayner (1982) used the moving window paradigm to compare reading when the word spaces were present to when they are replaced with an x. They found that saccade length is shorter when word space information is not available.

Boundary Study

The boundary study (Rayner, 1975) is another innovative paradigm that eye trackers and computers made possible. With the boundary study we can examine what information the reader is using inside the perceptual span (15 letters), but outside of the word that is being fixated. Figure 8 illustrates what the reader sees in this kind of study. While reading the words The old captain, the reader will be performing ordinary reading. When the reader reaches the word put, the key word of interest becomes available within the reader's fixation span. In this example the key word is ebovf. When the reader saccades from put to ebovf, the saccade will cross an invisible boundary which triggers a change in the text. Before the saccade finishes, the text will change to the correct text for the sentence, in this case chart. The reader will always fixate on the correct word for the sentence.

The critical word in this study is presented in different conditions including an identical control condition (chart), similar word shape and some letters in common (chovt), dissimilar word shape with some letters in common (chyft), and similar word shape with no letters in common (ebovf). The fixation times for the words both before and after the boundary are measured. The fixation times before the boundary are the same for the control condition and the three experimental conditions. After the boundary, readers were fastest reading with the control condition (chart), next fastest reading with similar word shape and some letters in common (chovt), third fastest with the condition with only some letters in common (chyft), and slowest with the condition with only similar word shape (ebovf). This demonstrates that letter information is being collected within the fixation span even when the entire word is not being recognized.



↑ **FIGURE 8** Figure 8: The string of letters ebovf after the boundary changes to chart during the saccade.

chart	Identical word (control)	210ms
chovt	Similar word shape Some letters in common	240ms
chyft	Dissimilar word shape Some letters in common	280ms
ebovf	Similar word shape No letters in common	300ms

↑ **FIGURE 9** Relative speed of boundary study conditions

Having letters in common played greater role in fixation times in this study. But it does not eliminate the role of word shape because of the combination of word shape and letters in common facilitates word recognition. Rayner (1975) further investigated what happens with a capitalized form of the critical word (CHART). This eliminates the role of word shape, but retains perfect letter information. They found that the fixation times are the same as the control condition! This demonstrates that it is not visual information about either word shape or even letter shape that is being retained from saccade to saccade, but rather abstracted information about which letters are coming up.

The eye movement literature demonstrates that we are using letter information to recognize words, as we are better able to read when more letters are available to us. We combine abstracted letter information across saccades to help facilitate word recognition, so it is

letter information that we are gathering in the periphery. And finally we are using word space information to program the location of our next saccade.

Evidence for Word Shape Revisited

So far I've presented evidence that supports the word recognition model, evidence that contradicts the serial word recognition model, and eye tracking data that contradicts the word shape model while supporting the parallel letter recognition model. In this section I will reexamine the data used to support the word shape model to see if it is incongruent with the parallel letter recognition model.

The strongest evidence for the word shape model is perhaps the word superiority effect which showed that letters can be more accurately recognized in the context of a word than in isolation, for example subjects are more accurate at recognizing D in the context of WORD than in the context of ORWD (Reicher, 1969). This supports word shape because subjects are able to quickly recognize the familiar word shape, and deduce the presence of letter information after the stimulus presentation has finished while the nonword can only be read letter by letter. McClelland & Johnson (1977) demonstrated that the reason for the word superiority effect wasn't the recognition of word shapes, but rather the existence of regular letter combinations. Pseudowords are not words in the English language, but have the phonetic regularity that make them easily pronounceable. Mave and rint are two examples of pseudowords. Because pseudowords do not have semantic content and have not been seen previously by the subjects, they should not have a familiar word shape. McClelland & Johnson found that letters are recognized faster in the context of pseudowords (mave) than in the context of nonwords (amve). This demonstrates that the word superiority effect is caused by regular letter combinations and not word shape.

The weakest evidence in support of word shape is that lowercase text is read faster than uppercase text. This is entirely a practice effect. Most readers spend the bulk of their time reading lowercase text and are therefore more proficient at it. When readers are forced to read large quantities of uppercase text, their reading speed will eventually increase to the rate of lowercase text. Even text oriented as if you were seeing it in a mirror will quickly increase in reading speed with practice (Kolers & Perkins, 1975).

Haber & Schindler (1981) found that readers were twice as likely to fail to notice a misspelling in a proofreading task when the misspelling was consistent with word shape (tesf, 13% missed) than when it is inconsistent with word shape (tesc, 7% missed). This is seemingly a convincing result until you realize that word shape and letter shape are confounded. The study compared errors that were consistent both in word and letter shape to errors that are inconsistent both in word and letter shape.

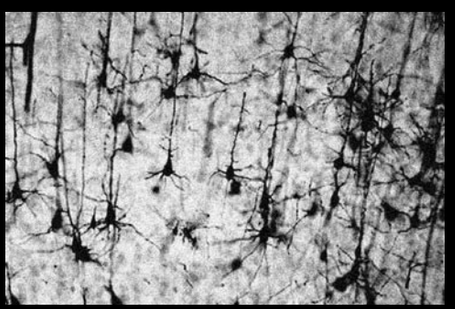
Paap, Newsome, & Noel (1984) determined the relative contribution of word shape and letter shape and found that the entire effect is driven by letter shape.
Figure 10 shows the example word than in each of the four permutations of same and different word shape, and same and different letter shape. As with Haber & Schindler, subjects fail to notice misspellings with the same word shape and same letter shape (tban, 15% missed) far more often than when there is a different word shape and letter shape (tman, 10% missed). The two in between conditions of different word shape with same letter shape (tnan, 19% missed) and same word shape with different letter shape (tdan, 8% missed) are illuminating. There is a statistically reliable difference between the larger number of proofreading errors when the letter shape is the same (tban and tnan) than when the letter shape is different (tdan and tman). While there is no statistically reliable difference between conditions with same word shape (tban and tdan) and different word shape (tnan and tman), more errors are missed when the word shape is different. This trend sharply contradicts the conclusions of the earlier studies.

than	SAME WORD SHAPE	DIFFERENT WORD SHAPE
SAME LETTER SHAPE	**tban** 15% missed	**tnan** 19% missed
DIFFERENT LETTER SHAPE	**tdan** 8% missed	**tman** 10% missed

↑ **FIGURE 10** Word shape and letter shape contributions to proofreading errors.

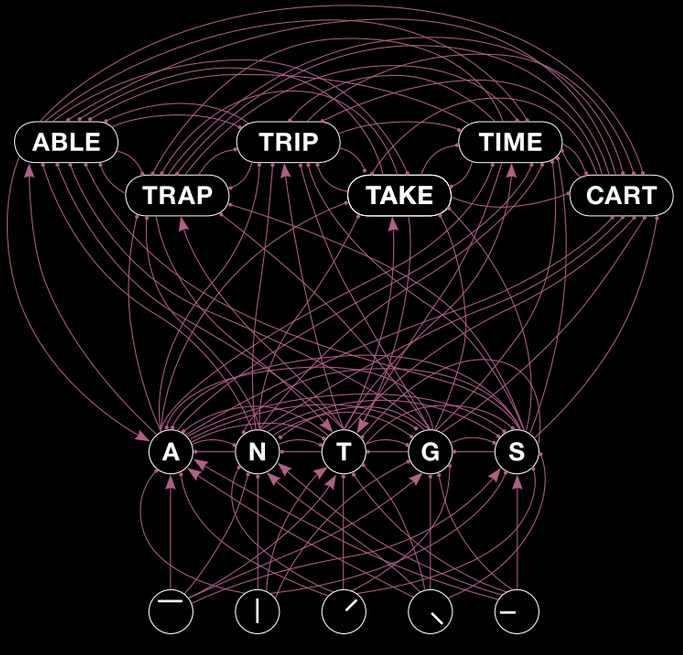
The final source of evidence supporting the word shape model is that text written in alternating case is read slower than either text in lowercase or uppercase. This supports the word shape model because subjects are able to quickly recognize the familiar pattern of a word written entirely in lowercase or uppercase, while words written in alternating case will have an entirely novel word shape. Adams (1979) showed that this is not the case by examining the effect of alternating case on words, which should have a familiar pattern when written in lowercase or uppercase words, and pseudowords, which should not have a familiar pattern in any form because the subjects would never have come across that sequence of letters before. Adams found that both words and pseudowords are equally hurt by alternating case. Since pseudowords are also impacted by alternating case, then the effect is not caused by word shape.
Further examination of the evidence used to support the word shape model has demonstrated that the case for the word shape model was not as strong as it seemed. The word superiority effect is caused by familiar letter sequences and not word shapes. Lowercase is faster than uppercase because of practice. Letter shape similarities rather than word shape similarities drive mistakes in the proofreading task. And pseudowords also suffer from decreased reading speed with alternating case text. All of these findings make more sense with the parallel letter recognition model of reading than the word shape model.
In the next section I will describe an active area of research within the parallel letter recognition model of reading. There are many models of reading within parallel letter recognition, but it is beyond the scope of this paper to discuss them all. Neural network modeling, sometimes called connectionist modeling or parallel distributed processing, has been particularly successful in advancing our understanding of reading processes.

↓ **FIGURE 11** A field of neurons and synapses in the cerebral cortex



Neural Network Modeling

In neural network modeling we use simple, low-level mechanisms that we know to exist in the brain in order to model complex, human behavior. Two of the core biological principles have been known for a long time. McCulloch & Pitts (1943, 1947) showed that neurons sum data from other neurons. Figure 11 shows a tiny two dimensional field of neurons (the dark triangles) and more importantly the many, many input and output connections for each neuron. Current estimates say that every neuron in the cerebral cortex has 4,000 synapses. Every synapse has a baseline rate of communication between neurons and can either increase that rate of communication to indicate an excitatory event or decrease the rate of communication to indicate an inhibitory event. When a neuron gets more excitatory information than inhibitory information, it will become active. The other core biological principle is that learning is based on the modification of synaptic connections (Hebb, 1949). When the information coming from a synapse is important the connection between the two neurons will become physically stronger, and when information from a synapse is less important the synapse will weaken or even die off.
The first well-known neural network model of reading was McClelland & Rumelhart's Interactive Activation model (1981). Figure 12 diagrams how this model works. The reader here is processing the letter T in the first position in a word. The flow of information here



↑ **FIGURE 12** McClelland & Rumelhart's Interactive Activation model: A few of the neighbors of the node for the letter T in the first position in a word, and their interconnections.

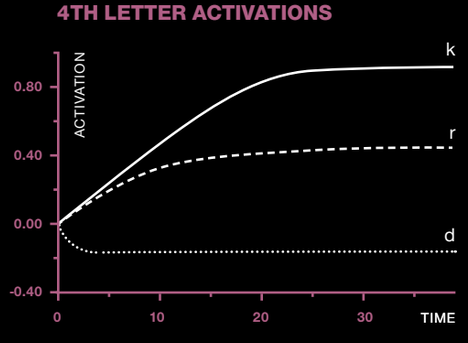
starts at the bottom where there are visual feature detectors. The two nodes on the left are active because they match the features of an uppercase T, while the three nodes on the right are not active because they don't match. Every node in the visual feature detector level is connected to every node in the letter detector level. The letters seen here apply only to the first letter of a word. The connections between the visual feature detector level and the letter level are all either excitatory (represented with an arrow at the end of the connection) or inhibitory (represented with a circle at the end of the connection). The letters A, T, and S all received some excitatory activation from the two left feature detectors because all three have a crossbar at the top of the letter (at least in this font). The inhibitory connections between each of the letters will result in the T being the most activated letter node because it has the most incoming excitatory activation. The letter node for T will then send excitatory activation to all the words that start with T and inhibitory activation to all the other words. As word nodes gain in activation, they will send inhibitory activation to all other words, excitatory activation back to letter nodes from letters in the word, and inhibitory activation to all other letter nodes. Letters in positions other than the first are needed in order to figure out which of the words that start with T is being read.

One of the joys of neural network modeling is that it's specific enough to be programmed into a computer and tested. The interactive activation model is able to explain human behaviors that it was not specifically designed for. For example when a human is shown the degraded stimulus in figure 13, it is very easy to figure out that WORK is the degraded word, but the computer simulation of this model can also solve this problem.

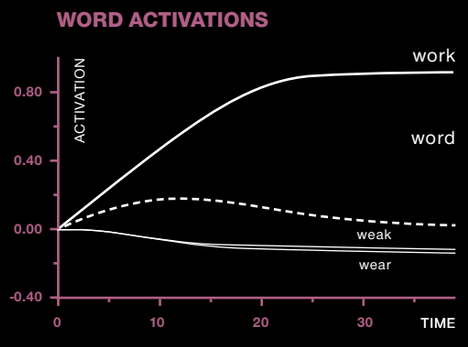
↑ **FIGURE 13** This degraded stimulus is easily read as WORK by human readers.

The computer simulation does not attempt to solve the visual perception problem, but rather is told which of the visual feature detectors are on for each letter position. For the fourth letter position the computer simulation is told that there is a vertical line on the left, a crossbar in the middle, and a diagonal pointing towards the bottom right. Figures 14

and 15 show the activation levels of certain letter and word nodes over time. Time in the computer is measured in epochs of activation events. Figure 14 shows the early activation equally rising for the k and r letter nodes. This is because the visual feature information supports both of those letters, while the d letter node is unsupported. During the early epochs the letter nodes are only receiving activation from the visual feature nodes, but later activation is provided by the word nodes. Figure 15 shows the activation among four words: work, word, weak, and wear. Since the first three letters of the word are not degraded, the letter nodes easily recognized them as w, o, and r for the first three positions respectively. These letters provide early activation for the words work and word, but not for weak and wear. The word nodes then start to send activation back down to the letter node level indicating that the fourth letter could be k or d. Since k is already an active letter node while d is an inactive node, the k node is further strengthened. This allows the k letter node and the word work to continuously increase in activation and send inhibitory activation to their competitors, the letter r and the word word. Similar activation patterns can also explain the word superiority effect.



↑ **FIGURE 14** The activation level over time for letter nodes in the fourth position of a word.



↑ **FIGURE 15** The activation level over time for four word nodes.

Seidenberg & McClelland (1989) and Plaut, McClelland, Seidenberg, & Patterson (1996) have made great progress in developing neural network models of reading that can account for more human reading behaviors. Both of these models concentrate on the reading processes that start after each of the letters in a word have been recognized. The internal representations for these models convert the letter information to phonemic information, which is seen as a mandatory step for word recognition. It is well known that words that have a consistent spelling to sound correspondence such as mint, tint, and hint are recognized faster than words that have an inconsistent spelling to sound correspondence such as pint (Glushko, 1979). These models are able to generate correct word pronunciations (i.e. read) without the use of specific word nodes. The more recent model is also able to read pseudowords at a near human rate and account for consistency and frequency effects. The Seidenberg & McClelland and Plaut et. al. models are able to simulate not only adult reading, but can also simulate a child learning to read. Initially the neural network model starts out with no knowledge about the relationship between letters and pronunciations, only that letters and sounds exist. The neural net goes through a training phase where the network is given examples of correct pronunciations for different words. After seeing a correct sample, the network will calculate the error in its guess of the pronunciation, and then modifies the strength of each of the nodes that are connected to it so that the error will be slightly less next time. This is analogous to what the brain does. After a few rounds of training, the model may be able to read a few of the most high frequency, regular words. After many rounds of training the model will be able to read not only words it has seen before, but words it hasn't seen before as well.

Conclusions

Given that all the reading research psychologists I know support some version of the parallel letter recognition model of reading, how is it that all the typographers I know say that we read by matching whole word shapes? It appears to be a grand misunderstanding. The paper by Bouma that is most frequently cited does not support a word shape model of reading. Bouma (1973) presented words and unpronounceable letter strings to subjects away from the fixation point and measured their ability to name the first and last letters. He found that:

A) Subjects are more successful at naming letters to the right of fixation than to the left of fixation.

B) When distance to the right of the fixation point is controlled, subjects are better able to recognize the last letter of a word than the first letter of word. This data explains why it is that we tend to fixate just to the left of the middle of a word.

Bouwhuis & Bouma (1979) extended the Bouma (1973) paper by not only finding the probability of recognizing the first and last letters of a word, but also the middle letters. They used this data to develop a model of word recognition based on the probability of recognizing each of the letters within a word. They conclude that 'word shape … might be satisfactorily described in terms of the letters in their positions.' This model of word recognition clearly influenced the McClelland & Rumelhart neural network model discussed earlier which also used letters in their positions to probabilistically recognize words. Word shape is no longer a viable model of word recognition. The bulk of scientific evidence says that we recognize a word's component letters, then use that visual information to recognize a word. In addition to perceptual information, we also use contextual information to help recognize words during ordinary reading, but that has no bearing on the word shape versus parallel letter recognition debate. It is hopefully clear that the readability and legibility of a typeface should not be evaluated on its ability to generate a good bouma shape.

Why I wrote this paper

I am a psychologist who has been working for Microsoft in different capacities since 1996. In 2000 I completed my PhD in cognitive psychology from the University of Texas at Austin studying word recognition and reading acquisition. I joined the ClearType team in 2002 to help get a better scientific understanding of the benefits of ClearType and other reading technologies with the goal of achieving a great on-screen reading experience.

During my first year with the team I gave a series of talks on relevant psychological topics, some of which instigated strong disagreement. At the crux of the disagreement was that the team believed that we recognized words by looking at the outline that goes around a whole word, while I believed that we recognize individual letters. In my young career as a reading psychologist I had never encountered a model of reading that used word shape as perceptual units, and knew of no psychologists who were working on such a model. But it turns out that the model had a very long history that I was unfamiliar with. ✖

KEVIN LARSON has been conducting experiments for over a decade towards understanding how we read and recognize words. He received his Ph.D. in cognitive psychology from the University of Texas in Austin. His current work in Microsoft's Advanced Reading Technologies group is focused on improving the quality of on-screen text.

References

Adams, M.J. (1979). Models of word recognition. *Cognitive Psychology*, 11, 133–176.

Bouma, H. (1973). Visual Interference in the Parafoveal Recognition of Initial and Final Letters of Words, *Vision Research*, 13, 762–782.

Bouwhuis, D. & Bouma, H. (1979). Visual word recognition of three letter words as derived from the recognition of the constituent letters, *Perception and Psychophysics*, 25, 12–22.

Cattell, J. (1886). The time taken up by cerebral operations. *Mind*, 11, 277–282, 524–538.

Fisher, D. F. (1975). Reading and visual search. *Memory and Cognition*, 3, 188–196.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 674–691.

Gough, P. B. (1972). One second of reading. In Kavanagh & Mattingly's *Language by ear and by eye*. Cambridge, MA: MIT Press.

Haber, R. N. & Schindler, R. M. (1981). Errors in proofreading: Evidence of syntactic control of letter processing? *Journal of Experimental Psychology: Human Perception and Performance*, 7, 573–579.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Mason, M. (1978). From print to sound in mature readers as a function of reader ability and two forms of orthographic regularity, *Memory and Cognition*, 6, 568-581.

Kolers, P. A. & Perkins, D. N. (1975). Spatial and ordinal components of form perception and literacy. *Cognitive Psychology*, 7, 228–267.

McClelland, J. L. & Johnson, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception and Psychophysics*, 22, 249–261.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375–407.

McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.

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

Meyer, D. E. & Gutschera, K. D. (1975). Orthographic versus phonemic processing of printed words. *Psychonomic Society Presentation*.

Monk, A. F. & Hulme, C. (1983). Errors in proofreading: Evidence for the use of word shape in word recognition. *Memory and Cognition*, 11, 16–23.

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

Pitts, W. & McCulloch, W. S. (1947). How we know universals: the perception of auditory and visual form. *Bulletin of Mathematical Biophysics*, 9: 127–147.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.

Pollatsek, A. & Rayner, K. (1982). Eye movement control in reading: The role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 817–833.

Pollatsek, A., Well, A. D., & Schindler, R. M. (1975). Effects of segmentation and expectancy on matching time for words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 328–338.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7, 65–81.

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology*, 12, 206–226.

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

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.

Smith, F. (1969). Familiarity of configuration vs. discriminability of features in the visual identification of words. *Psychonomic Science*, 14, 261–262.

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

Woodworth, R. S. (1938). *Experimental psychology*. New York; Holt.

Suggested Reading

If you're just looking for a couple of papers on reading psychology. I recommend these four:

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Review*, 124 (3), 372–422. This paper is an account of the eye movement field from the premier eye tracking researcher.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115. This is the most recent of the major neural network papers, and is available on David Plaut's website: http://www.cnbc.cmu.edu/~plaut/

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–407. This is one of the most cited reading papers of all time. If you are interested in reading acquisition this is the place to start.

Hoover, W. A. & Gough, P. B. (1990). The simple view of reading. *Reading & Writing*, 2(2), 127–160. This paper demonstrates that word recognition and context are two separate skills that are both necessary for reading.