

Mr. Chips: An Ideal-Observer Model of Reading

Gordon E. Legge, Timothy S. Klitz, and Bosco S. Tjan
University of Minnesota

The integration of visual, lexical, and oculomotor information is a critical part of reading. Mr. Chips is an ideal-observer model that combines these sources of information optimally to read simple texts in the minimum number of saccades. In the model, the concept of the visual span (the number of letters that can be identified in a single fixation) plays a key, unifying role. The behavior of the model provides a computational framework for reexamining the literature on human reading saccades. Emergent properties of the model, such as regressive saccades and an optimal-viewing position, suggest new interpretations of human behavior. Because Mr. Chip's "retina" can have any (one-dimensional) arrangement of high-resolution regions and scotomas, the model can simulate common visual disorders. Surprising saccade strategies are linked to the pattern of scotomas. For example, Mr. Chips sometimes plans a saccade that places a decisive letter in a scotoma. This article provides the first quantitative model of the effects of scotomas on reading.

Reading and Vision

A critical part of normal reading is the integration of visual data with lexical knowledge and eye-movement control. This integration process is the central computational problem addressed in this article. *Mr. Chips* is a computer program that implements an *ideal-observer principle* for this integration.

Damage to the early visual pathways can disrupt reading. More than 3 million Americans have *low vision*, defined as a visual disability, not correctable by glasses or contact lenses, that has an impact on daily function.¹ Most people with low vision have reading difficulties. A major goal of this article is to present a theoretical analysis of the role of vision in reading that encompasses both normal and low vision.

In a series of psychophysical studies, Legge and colleagues have investigated the effects on reading of many physical param-

eters of text and also the eye conditions of readers. They developed measures of reading speed that are sensitive to visual variables (Legge, Pelli, Rubin, & Schleske, 1985; Legge, Ross, Luebker, & LaMay, 1989). Text variables examined in their studies include character size (Legge, Pelli, et al., 1985; Legge, Ross, et al., 1989; Legge, Rubin, Pelli, & Schleske, 1985), number of characters visible in the field (Beckmann & Legge, 1996; Legge, Pelli, et al., 1985; Legge, Rubin, et al., 1985), contrast of letters (Legge, Rubin, & Luebker, 1987; Rubin & Legge, 1989), colors of letters and backgrounds (Legge, Parish, Luebker, & Wurm, 1990; Legge & Rubin, 1986), and letter fonts (Klitz, Mansfield, & Legge, 1995; Mansfield, Legge, & Bane, 1996).

This work was designed to understand the role of early visual processing in normal reading and to lay the foundation for understanding how eye disease affects reading. For example, normal reading speed is nearly unaffected by a reduction of text contrast from 100% to 10% (Legge et al., 1987). But some readers with low vision slow down for any reduction of text contrast. Reading deficits in readers with low vision caused by cataracts and other forms of cloudy ocular media can be explained by a simple model; these readers with low vision behave like normal readers preceded by a stage of contrast attenuation (Rubin & Legge, 1989).

Retinal disease, resulting in visual-field loss, has a major impact on reading speed (Legge, Ross, Isenberg, & LaMay, 1992; Legge, Rubin, et al., 1985). Age-related macular degeneration (AMD) is the leading cause of low vision and often results in a scotoma (blind spot) covering all or part of the central visual field. Other diseases, such as glaucoma and retinitis pigmentosa can produce peripheral field loss that spares only a small island of vision in the central retina. Yet other diseases,

Gordon E. Legge, Timothy S. Klitz, and Bosco S. Tjan, Minnesota Laboratory for Low-Vision Research, Department of Psychology, University of Minnesota. Bosco S. Tjan is now at Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany.

We thank Steve and Sharon Roe for suggesting the name *Mr. Chips* for the computer simulation, Steve Mansfield for computing the standard deviation of letter widths in several fonts, Paul Beckmann for help with graphics, and Andrew Luebker for help in solving programming bugs. We also thank Steve Mansfield, Paul Beckmann, Dan Kersten, and Irv Biederman for discussion of many aspects of ideal observers and reading.

This research was supported by National Institutes of Health Grants EY02934 and EY02857 from the National Eye Institute and HD07151 from the National Institute of Child Health and Human Development and U.S. Air Force Grant AFOSR 90-0274. Preliminary reports (Legge, 1992; Klitz & Legge, 1994) were given at meetings of the Association for Research in Vision and Ophthalmology, Sarasota, Florida, May 1992 and May 1994.

Correspondence concerning this article should be addressed to Gordon E. Legge, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455. Electronic mail may be sent via Internet to legge@eye.psych.umn.edu.

¹ *Low vision* is sometimes defined as a Snellen acuity with best correction of less than 20/60 (World Health Organization, 1966) and does not include individuals who are blind (approximately 200,000 in the United States) and who have no useful pattern vision.

like diabetic retinopathy, can produce a patchy pattern of scotomas across the visual field. Clinically, it is well-known that different patterns of scotomas produce different outcomes for reading. No theoretical framework exists for understanding these deficits. There have been attempts to train patients with central scotomas to use their peripheral vision (Goodrich & Quillman, 1977; McMahan, Hansen, Stelmack, Oliver, & Viana, 1993; Nilsson, 1990), but designers of these procedures have received little guidance from theories of reading. A major goal of our computational analysis is to establish a theoretical framework for understanding the effects of scotomas on reading.

Recently, there has been a growth of interest in reading tests among clinical vision researchers. It has become clear that single-letter acuity, the traditional measure of visual function (Snellen, 1864), is a poor predictor of performance in many real-world tasks including reading (Legge et al., 1992). Reading performance is becoming a de facto standard for assessment of functional vision. This has prompted the development of new tests of reading vision such as the Sloan M cards (Sloan, 1977), the Bailey-Lovie word-reading chart (Bailey & Lovie, 1980), the Pepper test (Baldasare, Watson, Whittaker, & Miller-Shaffer, 1986), and the MNREAD (a Minnesota reading test) Acuity Chart (Mansfield, Ahn, Legge, & Luebker, 1993).

The psychophysical work on visual factors in reading has not been well integrated with the literature on higher level perceptual and cognitive factors. This is because the psychophysical work has focused primarily on the role of low-level sensory factors in reading. The literature on higher level factors has usually taken the early sensory coding for granted. One reason is the finding that variations in reading speed among normally sighted participants are not closely tied to threshold differences in simple letter-recognition tasks (Jackson & McClelland, 1979). Such findings beg the question of the role of visual signals in reading and the impact of visual impairment on reading. More recent ideas about perceptual processes in reading—including properties of the visual span, the role of text information to the right of fixation, and the optimal viewing position within words—form the basis for bridging the gap between low-level and higher level visual encoding in reading. This work is reviewed later in comparison with the behavior of Mr. Chips. A major goal of this article is to integrate the findings from these two domains of research using a single theoretical framework.

Ideal Observers

An ideal observer is an algorithm that yields the best possible performance in a task that has a well-specified goal, usually in the face of visual noise or some source of uncertainty in the stimulus. The optimum performance level is dependent on the intrinsic information content of the stimuli and the task demands. The ideal observer provides an index of task-relevant information by showing the performance level that can be achieved when all of the information is used optimally. Comparison of human performance to ideal performance can establish whether human performance is limited by the information available in the stimulus or by information-processing limitations within the human.

The comparison of human and ideal observers in vision is

similar to Chomsky's distinction between competence and performance in linguistic theory (Chomsky, 1965). An ideal observer embodies the "competence" associated with a particular visual task. Human observers have access to the same information but may be limited in their performance by a variety of perceptual or cognitive factors. In relation to Marr's (1982) three levels of analysis (computational theory, algorithm, and hardware implementation), an ideal observer contains an algorithm that takes into account all aspects of the computational theory relevant for maximizing performance.

Anderson has advocated the rational analysis of cognition (cf. Anderson, 1991) in which basic cognitive functions are explained as optimal adaptations to the statistical structure of information in the environment. Anderson's approach shifts the emphasis in studies of cognition from properties of mechanisms to the informational constraints facing the participant and has the same spirit as the ideal-observer approach.

Hecht, Shlaer, and Pirenne (1942) introduced the ideal observer to vision with their work on quantal limits on light detection. Subsequently, ideal observers have been invoked in studying many simple detection and discrimination tasks, culminating in the elegant sequential ideal-observer analysis of Geisler (1989). Ideal-observer analysis has also been used in studying more complex visual-information processing tasks, including detection of mirror symmetry (Barlow & Reeves, 1979), discrimination of dot density (Barlow, 1978), discrimination of the number of dots in displays (Burgess & Barlow, 1983), estimation of statistical parameters in graphical displays (Legge, Gu, & Luebker, 1989), letter recognition (Parish & Sperling, 1991; Pelli, Burns, Farrell, & Moore, in press; Solomon & Pelli, 1994), word recognition (Pelli, Farrell, & Moore, 1995) and object recognition (Braje, Tjan, & Legge, 1995; Liu, Knill, & Kersten, 1995; Tjan, Braje, Legge, & Kersten, 1995).

In vision research, ideal observers have usually taken the form of a Bayesian observer that makes the best inference from the image data (Kersten, 1990). Mr. Chips uses a different optimizing principle—*entropy minimization*—where entropy is defined in information theory (Shannon & Weaver, 1949). Mr. Chips uses all available visual, lexical, and oculomotor information to select a *saccade* (rapid eye movement) that minimizes uncertainty (i.e., entropy) about a word in the text. Attneave (1954) proposed that concepts from information theory could be useful in understanding visual pattern recognition.

The ideal-observer approach has several advantages: (a) Ideal observers can be used to quantify the information available in a task; (b) ideal performance provides a benchmark for evaluating human performance; (c) the behavior of ideal observers can shed light on corresponding human behavior: we show that Mr. Chips exhibits regressive saccades, word skipping, and an optimal viewing position in words, similar to human readers; (d) using ideal performance as a benchmark, we can evaluate simpler, suboptimal strategies: For example, we show that Mr. Chips's ideal algorithm can be nearly matched by simple heuristics that require almost no computation; and (e) ideal-observer analysis forces explicit attention to the underlying computational theory.

Ideal-observer analysis has two disadvantages that are also shared by Mr. Chips. First, specification of all assumptions and

sources of information often means working in a simplified world. Mr. Chips performs a very simple reading task and is subject to several simplifying assumptions. Secondly, even in simple contexts, ideal observers can be hard to derive mathematically or difficult to implement as a computer program. Often, they require an inordinate number of cycles by a central processing unit.

Scope of This Article

The focus of this article is on the integration of visual, lexical and oculomotor information. Mr. Chips accepts visual input in the form of recognized (or partially recognized) letters and is not concerned directly with the machinery of letter recognition. Mr. Chips stops short of including syntactic, semantic and higher level constraints. We describe how the effect of context on the model's performance can be approximated by manipulating the size of the available lexicon.

In the Method: General section, we provide an overview of the model and its implementation as a computer simulation. Following that, there is a detailed description of the algorithm embodied in Mr. Chips. The results are presented in three sections dealing with the three types of information used by Mr. Chips. The results from the simulations are presented in relation to our current understanding of normal and low-vision reading. The article concludes with a discussion of possible theoretical extensions of Mr. Chips.

This article uses methods and concepts from visual perception, clinical-vision research, and artificial intelligence. This interdisciplinary approach can shed light on the role of vision in a task of fundamental importance to participation in modern society: reading.

Method: General

Overview of the Ideal Reader

The *ideal reader* is a model that specifies how three sources of information—visual, lexical, and saccadic—are combined optimally for recognizing words and planning saccades. Mr. Chips is the name² of a computer simulation of the ideal reader. Some simplifications were adopted in the software implementation and are discussed later. The key elements of the ideal-observer model of reading are depicted in Figure 1.

The *text* is constructed by drawing words at random from a known lexicon according to a probability distribution and stringing them together with spaces between them. All the words lie on a single long line (Mr. Chips does not have to worry about return saccades at the ends of lines.)

Mr. Chips's *task* is to read the text in the minimum number of saccades, identifying the words in order without error.

Mr. Chips obtains *visual data* by sampling the text through a "retina" with regions of three kinds: (a) high-resolution vision in which letters can be identified, (b) low-resolution vision (relative scotomas) in which spaces can be distinguished from letters but letters cannot be identified, and (c) blind spots (absolute scotomas) in which there is no vision. An example of a "normal" retina is illustrated in Figure 2 (discussed in the next

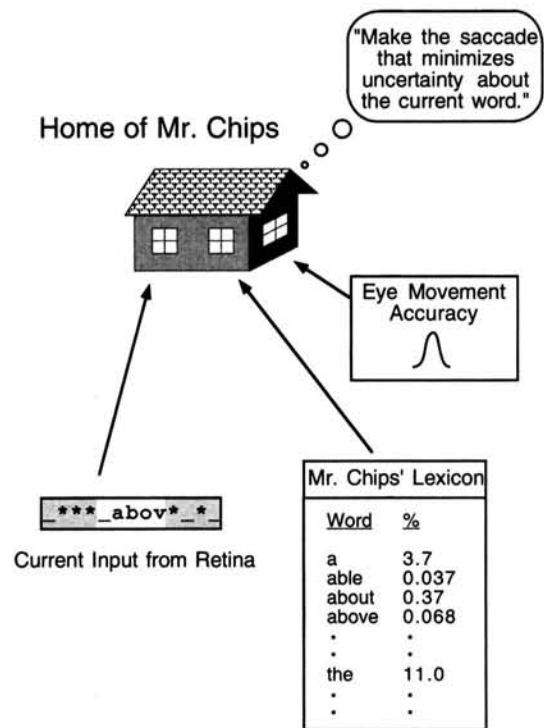


Figure 1. Cartoon of the ideal reader, illustrating the sources of information available and the optimizing principle.

section). "Abnormal" retinas (simulating low vision) consist of unusual patterns of high- and low-resolution zones, such as the retina with a central scotoma in Figure 3. A major purpose of this research is to study the performance of the ideal reader when different types of retinal sampling are used. (Extension of the model to allow finer gradations of partial information through the retinal slots is considered in the Discussion section.)

Mr. Chips's *lexical knowledge* consists of the list of words, and their relative frequencies, used to generate the text. Any arbitrary lexicon is possible. Most of our examples used a lexicon composed of the 542 most common words in written English, excluding words with uppercase letters or punctuation (Carroll, Davies, & Richman, 1971). Smaller lexicons were produced by subsampling the original lexicon. The reduced uncertainty associated with small lexicons provided a method for simulating context effects.

Mr. Chips's *motor knowledge* consists of statistical knowledge of the accuracy of his saccades. When he plans a saccade, he knows the parameters (mean [M] and standard deviation

² The name *Mr. Chips* is taken from *Goodbye, Mr. Chips* (Hilton, 1934). This novel tells the story of an English schoolmaster. Although Mr. Chips of literature had a modest intellect, he became the fount of wisdom and knowledge within the small world of the boarding school where he lived and taught throughout his career: "They all asked him questions, as if he were some kind of prophet and encyclopedia combined" (Hilton, 1934, chap. 16). Mr. Chips, the computer simulation, is omniscient in the restricted domain of a very simple reading task.

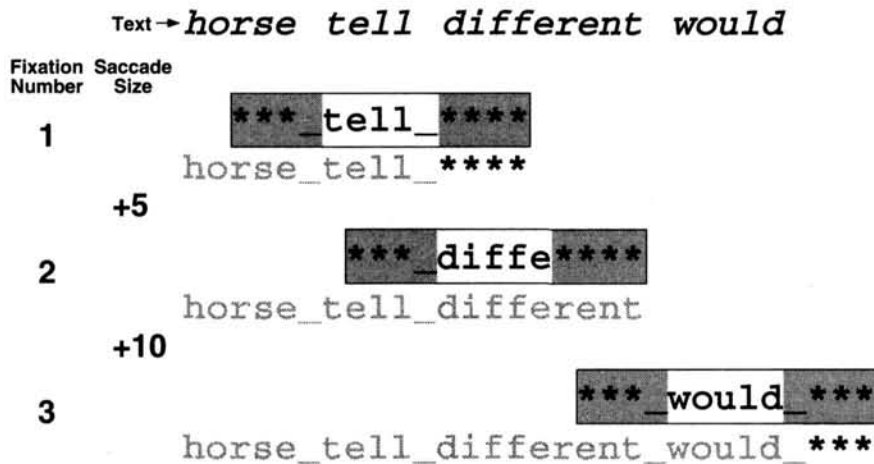


Figure 2. Series of three fixations by Mr. Chips. A portion of text is shown in italics at the top. Mr. Chips's retina contains a high-resolution central region with a span of five letters in which letters are recognized. The central region is flanked by low-resolution peripheral regions, shown in gray, each four characters wide. The asterisks (*) in the gray regions indicate that letters have been detected but not recognized. The underlined space () indicates that a space has been recognized. The text lying immediately below the retina on each fixation shows Mr. Chips's current state of knowledge. The word in bold indicates the one Mr. Chips is trying to figure out.

[SD]) of the probability distribution of landing sites. As a simplifying assumption, we considered only Gaussian error.

Mr. Chips operates according to an *entropy-minimization principle*. He identifies words as he moves from left to right through the text. At any point, Mr. Chips tries to identify the *current word*, that is, the leftmost word in the text that has not been identified already. Often, Mr. Chips has partial information about the current word (some letters known or word length known). Mr. Chips finds the subset of words from the lexicon that is consistent with the partial information. He then decides on his next saccade according to the following rule:

Make the saccade that minimizes uncertainty (as defined in information theory) about the current word. If there is a tie, choose the saccade that moves the retina furthest to the right.

Mr. Chips cannot leave the current word until he has identified it unambiguously (i.e., its uncertainty is zero, or less than some specified entropy tolerance). In the Discussion, we consider generalizations of the model in which the ideal observer analyzes more than one word at a time and plans more than one saccade at a time.

An Example

Figure 2 depicts a series of three fixations by Mr. Chips. A portion of the text is shown in italics at the top. In this example, Mr. Chips's retina has a high-resolution central region flanked on either side by low-resolution peripheral regions. The low-resolution regions are depicted in gray on the figure.

On Fixation 1, Mr. Chips sees "tell" in his central vision, in which the underscore () stands for the recognition of a space. The asterisks (*) in the gray regions indicate that letters

have been detected in the low-resolution peripheral regions. The text lying immediately below the retina is "horse tell ****" and shows Mr. Chips's current state of knowledge. The **** symbols at the right are in bold to show that this is the word Mr. Chips has to figure out next. From our bird's eye view, we know the word is *different*. Mr. Chips knows only that the next word contains at least four letters.

According to his entropy minimization calculation (see the section on the Method: Mr. Chips's Algorithm), he minimizes expected uncertainty about the unknown word by moving his retina five letter positions to the right. This has the effect of left-justifying the word in his high-resolution central vision. (Consideration of saccade errors is deferred until a later section.)

On Fixation 2, the retina shows "##_diffe####," thus Mr. Chips knows that the word begins *diffe* and has at least nine letters. He uses lexical inference; the only word in his lexicon meeting these constraints is *different*. This is why the text just below the retina on Fixation 2 shows the full word "different."

What saccade should Mr. Chips make next? He knows without seeing it that a space follows *different*, and the next unknown word begins after the space. He minimizes expected uncertainty about that word by making a long saccade of 10 spaces over the unseen end of *different* to bring *would* into central vision on Fixation 3.

Mr. Chips and Reading Speed

Reading speed is usually measured in words per minute and is a common measure of reading performance (Carver, 1990; Legge, Pelli, et al., 1985; Tinker, 1963). One problem with this measure is that mean word length varies from passage to pas-

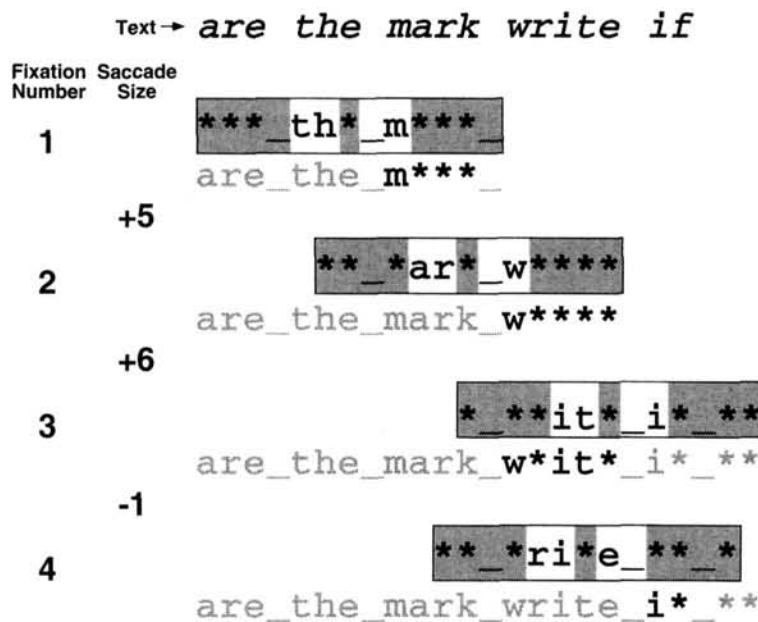


Figure 3. Series of four fixations by Mr. Chips. His retina has a one-character central scotoma (i.e., a slot in which a letter cannot be recognized, but a letter can be distinguished from a space). Notice that the third saccade is a regression (leftward saccade). Regressions are present in the behavior of an ideal reader. A sample animation for Figure 3, demonstrating a sequence of Mr. Chips's saccades, is available at the following WorldWideWeb address: <http://vision.psych.umn.edu/chips.html>. * = letters have been detected but not recognized; _ = a space has been recognized.

sage, increasing with text difficulty. Carver (1990) has shown that this problem can be remedied by measuring reading speed in "standard-length words" (p. 9) per minute, where each six characters counts as one standard-length word. Carver has shown that a participant's reading speed is constant in standard-length words per minute across text difficulty, provided the grade level of the text does not exceed the reading level of the participant. He has acknowledged the obvious fact that reading speed decreases when text difficulty equals or exceeds the grade level of the reader.

Carver's (1990) definition of reading speed is equivalent to measuring speed in characters per second. The link to eye-movement parameters is straightforward: Reading speed is equal to the mean saccade length (characters per saccade) divided by the mean time between saccades (seconds per saccade). In human reading, most of the time between saccades is taken up by fixations (roughly 250 ms), with relatively little time taken by the saccade itself (roughly 25 ms).

The computational theory embodied in Mr. Chips does not take time variables into account. The model's value of mean saccade length can be converted to reading speed in characters per second by assuming a mean value for the time between saccades.

To include time explicitly in the computational theory would require detailed specification of the temporal constraints governing visual data acquisition, lexical access, and saccade planning. This extended theory lies beyond the scope of this article.

In short, for Mr. Chips, reading speed is equivalent to mean

saccade size. Mr. Chips's computational theory does not incorporate explicit consideration of time variables.

Software Implementation

The simulation was written in Turbo-C (1987) and executed on several IBM-compatible computers. All of the simulation results cited in this article were obtained from the same version of the software (chips43; Legge, 1995). Histograms and summary statistics are based on 2,000-word texts.

Method: Mr. Chips's Algorithm

This section describes the computational theory and algorithm. The results in the following sections can be understood without mastery of these details.

In the computer simulation, the retina can be any linear arrangement of (a) high-resolution slots in which letters are identifiable, and (b) low-resolution slots (relative scotomas) in which a space can be discriminated from a letter, but letters cannot be discriminated from each other. Blind spots (absolute scotomas) are not allowable in the simulation because of the multiple-segmentation problem they introduce (discussed later).

The text consists of a sequence of words drawn at random from the lexicon. The ideal reader, Mr. Chips, proceeds from left to right through the text, identifying the words in order.

Uncertainty About the Current Word

In general, Mr. Chips has identified the first n words in the text, but he is uncertain about the $n + 1$ word. Refer to this word as the current word. The retina may provide two types of partial information about the current word that reduces the set of lexical alternatives: word length and the identity of one or more letters.

If the retinal input contains the bounding space to the right of the current word, the exact length of the current word is known. Otherwise, Mr. Chips knows only a lower bound on word length. If the word length is known to be four characters, the number of possible words is reduced from 542 (entire lexicon) to 164 (number of four-letter words in the lexicon). If the word length is known to be four or more, there are 437 possibilities.

The retina³ may yield the identity of one or more of the letters in the current word (not necessarily the leading letters). For example, suppose that only the leading letter c is known. The standard lexicon contains 31 words beginning with c . If Mr. Chips knows that the current word begins with c and is four letters long, there are only 5 possible words—*call*, *came*, *city*, *cold*, and *come*.

If the current word lies beyond the right margin of the retina (e.g., Fixation 2 in Figure 2), there is no partial information and any of the 542 lexical entries is possible.

Mr. Chips constructs a temporary table of all lexical entries consistent with the partial information about the current word. Suppose there are k words in this table, W_1, W_2, \dots, W_k , with corresponding probabilities in the lexicon of P_1, P_2, \dots, P_k . If $k = 1$, the current word has been identified, and the next word in the text string becomes the current word. If $k > 1$, Mr. Chips remains uncertain about the current word. The probability of each of the k words, conditional on the partial information, is designated p_i :

$$p_i = \frac{P_i}{\sum_j P_j} \quad (1)$$

Table 1 gives values for four-letter words beginning with c in the standard lexicon.

The entropy H associated with the set of candidate words is

$$H = -\sum_i p_i \log(p_i) \quad (2)$$

Table 1
Five Four-Letter Words Beginning With c in
the Standard Lexicon

Word	Absolute probability	Conditional probability
<i>call</i>	.00041	.097
<i>came</i>	.00145	.344
<i>city</i>	.00055	.130
<i>cold</i>	.00043	.102
<i>come</i>	.00138	.327

where p_i is the relative probability of the i th word in the list. The entropy of the five candidate words in Table 1 is 2.10 bits.

Saccade Planning in the Absence of Noise

Mr. Chips's goal is to select the saccade that minimizes his uncertainty about the current word. He needs to compute the expected uncertainty following each possible saccade length L . In this analysis, saccade lengths are assumed to be an integer number of character positions.

First, Mr. Chips must calculate the range of saccade sizes that could yield useful information. These include all saccades that leave at least one slot of the retina overlapping at least one letter of the current word. The most negative value of L (largest possible regression) is determined from the known position of the leading letter in the current word. Call the leftmost saccade length L_{\min} . From the list of candidate words in the temporary table, he knows the rightmost position of the final letter in the longest possible word, and this determines the largest rightward saccade. Call this value L_{\max} .

Following a saccade of length L in the range from L_{\min} to L_{\max} , the retina may yield additional visual information about the current word. The amount of new information (entropy reduction) depends on which of the k words is actually present. Mr. Chips must consider all k possibilities. From his temporary table, he first considers word W_1 with relative probability p_1 (in Table 1, *call* with probability .097). Given a saccade of length L and the actual presence of the word W_1 , the retinal data is consistent with a subset of the original k words. Suppose there are k' of these words ($k' \leq k$). Mr. Chips creates a sub-table of these k' words and conditional probabilities p'_i , analogous to the original temporary table. From this subtable, he computes a conditional entropy (i.e., the remaining entropy of the current word after a saccade of length L , assuming the actual presence of word W_1):

$$H(L, W_1) = -\sum_{i=1}^{k'} p'_i \log(p'_i) \quad (3)$$

Mr. Chips repeats this entropy calculation for all k words in the original temporary table. The expected entropy $H(L)$ associated with a saccade of length L is the probability-weighted average of the conditional entropies:

$$H(L) = \sum_{i=1}^k p_i H(L, W_i) \quad (4)$$

In other words, for each of the k candidate words in the temporary table, Mr. Chips computes the uncertainty he would have about the current word after a saccade of length L . He averages over all k words to get the expected uncertainty associated with a saccade of length L .

In our example, Mr. Chips's temporary table (Table 1) contains five possible words. Suppose that a saccade of length L reveals the second letter of the current word but no other infor-

³ Visual input concerning the current word may come from more than one fixation. Mr. Chips accumulates information across fixations.

mation. If the actual word is *call*, he identifies *a*, and reduces the set of candidates to *call* or *came*. If the actual word is *came*, his state of knowledge is the same. From his knowledge of the absolute probabilities of *call* and *came* in the lexicon, he computes their relative probabilities to be .22 and .78. From these values, he computes that $H(L, call) = H(L, came) = 0.76$ bits. If the actual word is *city*, knowledge of the second letter reduces uncertainty to zero. If the second letter is *o*, the set of candidates is reduced to *cold* and *come*. The corresponding entropy values are $H(L, cold) = H(L, come) = 0.79$ bits. The expected uncertainty associated with a saccade L is the sum of the five conditional entropies weighted by the relative probabilities of the five words: $H(L) = .097 \times 0.76 + .344 \times 0.76 + 0 \times 0.130 + .102 \times 0.79 + .327 \times 0.79 = 0.674$ bits.

Mr. Chips computes $H(L)$ for all saccade lengths L in the range L_{min} to L_{max} . He looks for the minimum value of $H(L)$. If there is a tie, he selects the saccade L that advances him furthest to the right.

Once the saccade is actually executed, the new visual input may be sufficient to unambiguously identify the current word. If so, Mr. Chips turns his entropy analysis to the next word. If uncertainty remains concerning the word, Mr. Chips runs through the entropy analysis again, taking his temporary table of k candidate words as the starting point.⁴ He uses the new visual data to winnow down the set of candidate words.

Saccade Planning in the Presence of Noise

Deterministic saccade errors, such as systematic undershoots, pose no problems for Mr. Chips because he can compensate for them. For example, if he knows that his saccade controller signals $L - 1$ when he asks for a saccade of length L , he asks for $L + 1$ in order to get L .

Suppose there is random error in the saccade system so that there is a distribution of possible landing sites following a saccade of an intended length. Mr. Chips is confronted with two new problems: fixation linking and entropy averaging.

Fixation linking. If the saccade length is not deterministic, Mr. Chips must figure out where his retina has landed. This computation is not a part of the current algorithm, so we assume that Mr. Chips always knows his retinal position relative to the text. We return to fixation linking in the Discussion.

Averaging entropy across landing sites. We assume a Gaussian distribution of saccade landing sites with mean equal to the intended saccade length L and standard deviation S . According to Fitts' law (Fitts, 1954), S is proportional to L . A Weber fraction S/L of 5% to 10% may be typical of human saccades under high-precision conditions (Kowler and Blaser, 1995).

We convert the Gaussian noise into discrete probabilities for saccades of integer length. We call these distributions *landing-site probability distributions*.⁵ For example, for a Gaussian distribution of mean saccade length 7 and standard deviation 1.0, the area under the normal distribution curve from 6.5 to 7.5 is 0.3833. This is taken as the probability of a saccade of length 7. Similarly, the area from 5.5 to 6.5 is 0.2416, and this is the probability of a saccade of 6, etc.

Let the landing-site probabilities associated with a saccade

of intended length L be designated $P_L(x)$, in which x is the saccade length. Because Mr. Chips knows the mean and standard deviation of his saccade noise, he can construct the landing-site probability distribution.

For an intended saccade L , the entropy averaged across the possible landing sites is H_L ⁶:

$$H_L = \sum_x P_L(x) H(x). \quad (5)$$

H_L is the expected uncertainty about the current word, following a saccade of intended length L , but taking saccade noise into account.

When saccade noise is present, Mr. Chips computes the values of H_L for saccades in the range L_{min} to L_{max} . He selects the saccade with minimum value of H_L , taking the longest rightward saccade when there are ties.

Computational Complexity

The number of entropy calculations increases approximately as k^2 , where k is the number of candidate words in the temporary table. The values of k are large when Mr. Chips has only word-length information. Under these conditions, the execution time on the IBM-compatible computers was prohibitively long (e.g., several minutes to compute the next saccade). To deal with this problem, Mr. Chips stores off-line results for the cases in which only word-length information is available. For the standard lexicon, where word lengths range from 1 to 10 letters, there are 19 possible states of knowledge concerning word length: 1 letter long, at least 1 letter long, 2 letters long, at least 2 letters long, . . . , 10 letters long. Mr. Chips does an off-line analysis of the entropies $H(L)$ associated with each of these 19 cases, and each possible state of overlap between the retina and target word. The resulting entropies are stored in a file. Then, when 1 of these cases occurs, Mr. Chips refers to the file instead of crunching through all the burdensome computation.

This type of "table look-up" is fast and practical when only word-length information is available. But, table look-up is impractical when the partial information about the current word contains one or more identified letters. There are too many possible cases to store, so the entropy calculations are done on-the-fly.

Results: Visual Analysis

Regressions: An Emergent Phenomenon

Figure 3 shows a sequence of four fixations when Mr. Chips has a retina with a central scotoma. The scotoma is one letter wide.

⁴ The model can be given a nonzero uncertainty criterion for moving to the next word. In this case, Mr. Chips occasionally misidentifies words, the error rate increasing as the uncertainty criterion rises.

⁵ We truncate the Gaussian so that we need not be concerned with highly improbable cases. For example, 99.93% of cases fall within 3.50 standard deviations.

⁶ The range of the summation over x is limited by truncation of the Gaussian. In any case, there is no need to include saccades outside the range L_{min} to L_{max} because there is no entropy reduction.

Fixation 1. Although the *e* of the lies in the scotoma, Mr. Chips knows the word because there is no other three-letter word beginning *th* in his lexicon. The next unknown word begins with *m* and has four letters. Mr. Chips makes a five-letter saccade. (Consideration of saccade noise is deferred until a later section.)

Fixation 2. Mr. Chips remembers *m* from the previous fixation and sees "ar." The only consistent four-letter word in his lexicon is *mark*. He begins working on the next word. He knows that it begins with *w* and is at least five letters long. He makes a six-letter saccade.

Fixation 3. Mr. Chips sees the third and fourth letters of the word in the left two high-resolution slots of his retina. He now knows that the word has the form "w*it*." There are two candidate words in his lexicon: *write* and *white*. The only way of resolving the ambiguity is to back up to identify the second letter in the word. He is compelled to make a one-character regression.

Fixation 4. Following the regression, he learns that the critical letter is *r*.

Why did Mr. Chips plan a saccade that overshot the mark? This behavior was a calculated gamble. The six-letter saccade would yield, on average, the most information about this word. In the case at hand, he was unlucky and needed to make a regression to resolve the remaining ambiguity.

The existence of regressive saccades in the behavior of the ideal reader is an example of an emergent phenomenon. Often considered a sign of poor reading, regressions occur for the ideal reader as a consequence of the underlying optimizing principle. Regressions can be ideal.

The proportion of regressions by Mr. Chips depends on his retinal structure and the nature of the saccade noise. Table 2 shows sample percentages. The number of regressions rises when there is saccade noise or when scotomas are present.

The proportion of regressions in normal human reading depends on the complexity of the text, ranging from about 3% for light fiction to 18% for biology text (Rayner & Pollatsek, 1989, chap. 4). An increased proportion of regressions occurs for normally sighted poor readers (Just & Carpenter, 1987) and also for patients with central scotomas (Bullimore & Bailey, 1995).

It is sometimes assumed that the presence of many regressions in low vision is a sign of poor reading strategy and that training should yield a more orderly sequence of forward saccades. The behavior of Mr. Chips indicates, however, that a higher rate of regressions would be expected for patients with central scotomas if they adopt an optimal reading strategy.

Visual Span

We refer to the *visual span* in reading as the number of characters that can be recognized on each glance. Our use of the term visual span is similar to O'Regan's (1990, 1991) usage. He defined the visual span as the distance on either side of the fixation point within which characters of a given size could be recognized. Because letters flanked by other letters are more difficult to identify in peripheral vision (cf. Bouma, 1970), the visual span for reading is smaller than the visual span for isolated letters.

The notion of visual span differs from the concept of *perceptual span* (McConkie & Rayner, 1975; Rayner & McConkie, 1976). Perceptual span is defined in terms of the functional demands of reading and includes factors in addition to letter recognition (Rayner & Pollatsek, 1989, chap. 4). Operationally, it refers to the region of visual field that influences eye movements and fixation times in reading. Rayner and McConkie (1976) estimated that the perceptual span extends 15 characters to the right of fixation and 4 characters to the left.

Mr. Chips's visual span is the number of high-resolution slots in central vision. His perceptual span also includes his low-resolution peripheral regions, to the extent that they influence his reading behavior.

Figure 4 shows histograms of saccade sizes when Mr. Chips had normal retinas with visual spans ranging from 1 to 11. Dark bars show results in the absence of noise and gray bars in the presence of saccade noise. In the noise-free case, the mean saccade lengths M increase linearly with the size of the visual span N according to the rule:

$$M = N + 1. \quad (6)$$

The curve labeled *Mr. Chips* in Figure 12A plots this relation.

How wide is the visual span in normal human reading? O'Regan (1990, 1991) presented a theoretical model of the human visual span. According to the model, the number of adjacent letters recognizable in central vision is determined by the size of the critical features in the letters, the falloff in the eye's spatial resolution away from the fixation point, and the geometry of the display surface. His model predicts a visual span of about 15 for letters subtending 0.4° .⁷ For magnified letters subtending 6° , the model's visual span drops to about 10 letters.

O'Regan, Levy-Schoen, and Jacobs (1983) measured the recognition of letters (flanked by numerals) as a function of their retinal eccentricity. They defined visual span in terms of the eccentricity within which letters could be recognized above

Table 2
Percentages of Regressions by Mr. Chips

	Normal retina	Central scotoma	Alcatraz
Noise condition			
No noise	<1%	<1%	1.7%
Noise present	3.8%	12%	23%

Note. The saccade noise was drawn from a Gaussian distribution with a standard deviation of 1.0, independent of saccade size.

⁷ O'Regan's (1991) theoretical equation for the number of letters, N , in the visual span (full width) is: $N = 2[(G/R_0MW) - (D/MW)]$, where G is the "grain" size (critical feature size for the letters in question), R_0 is the resolution limit of the critical features at the center of the fovea, M is an eccentricity-scaling factor similar to M scaling, W is the size of the letters, and D is the viewing distance. The visual span N is equal to 15.28 letters for the following values, suggested by O'Regan: $D = 40$ cm, $W = 0.3$ cm (equivalent to letters subtending 0.42°), $G = .08$ cm; $R_0 = 0.00029$ rad (equivalent to a resolution limit of 1.0 min-arc for letters at the center of the fovea), and $M = 100/\text{rad}$.

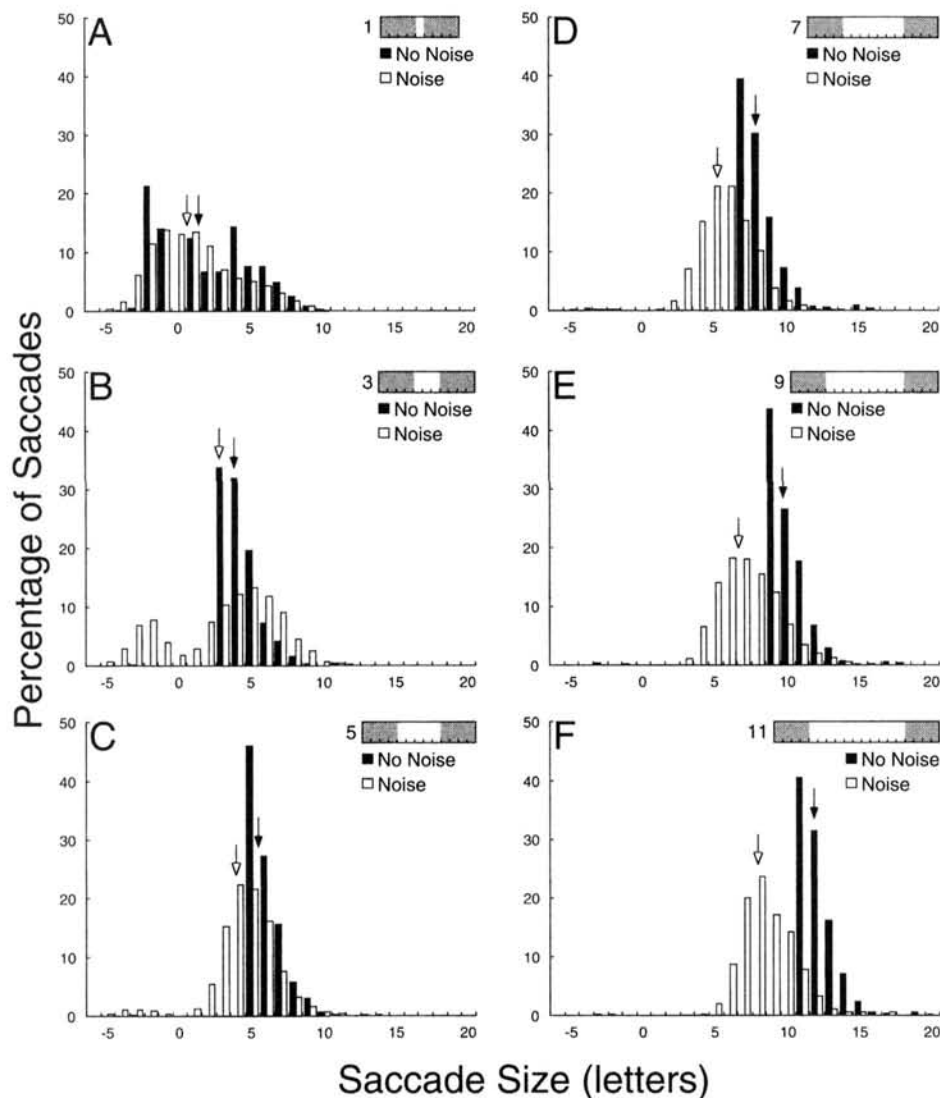


Figure 4. Saccade histograms for retinas with visual spans of 1, 3, 5, 7, 9, and 11 characters. Dark bars show values in the absence of saccade noise, and open bars show values in the presence of noise. The noise had a standard deviation of 1.0, independent of saccade length. Arrows point at the mean values for the two distributions. Each distribution was based on a text of 2,000 words.

some criterion level. For criteria of 50% and 90% correct, the visual spans were 22 letters and 10 letters respectively.

Rayner and Bertera (1979) used an eye-tracking method to mask letters (each subtending about 0.33°) surrounding the point of fixation during reading. When the mask covered the central 7 letters, reading speed was very low, about 12 words per minute. When the mask covered 11 letters, reading was essentially impossible. These results imply that human readers have a visual span of 7 to 11 letters.

In a recent study, Legge, Ahn, Klitz, and Luebker (1995) measured reading times as a function of word length using a rapid-serial-visual-presentation (RSVP) method. From their data, they estimated the size of the visual span for high-contrast 1° and 6° letters to be 10.6 and 5.3 letters respectively.

These empirical estimates point to a visual span in human reading of 9 or 10 letters for normal text ($\sim 0.3^\circ$ – 1.0°) and somewhat smaller for magnified text. With a visual span of 9, Mr. Chips can achieve a mean saccade length of about 10 letters in the absence of noise and 7 letters in the presence of saccade noise (Figure 4). Estimates of mean saccade size in normal human reading range from about 7 to 9 characters (cf. Rayner & Pollatsek, 1989, chap. 4).

Figure 4 shows that the proportion of regressions is very small for a visual span of 9 ($<1\%$). Humans exhibit a higher proportion of regressions when reading real text. Some of their extra regressions may be a consequence of nonideal behavior (e.g., regressions following saccadic overshoots). Others may be related to semantic or syntactic constructions unknown by

Mr. Chips, such as regressions to the referent of a pronoun (Carpenter & Just, 1977).

There is evidence that mean saccade size in human reading, measured in characters, is constant over a range of small character sizes: 0.35° – 0.70° (Morrison & Rayner, 1981),⁸ and 0.12° – 0.48° (O'Regan et al., 1983). O'Regan et al. argued for a decoupling between visual span and saccade size because they estimated the visual span to change by 30% over the range of character sizes for which mean saccade size remained constant. Although factors besides visual span surely play a role in determining saccade size, the results from Mr. Chips strongly suggest some linkage when the visual span changes by a large amount.

Shrinkage of the visual span may play an important role in explaining reduced reading speed in low vision. Most people with low vision require magnified text. Theoretical and empirical estimates agree that the visual span decreases for highly magnified characters (6° characters). Shrinkage of the visual span may also result from reduced retinal-image contrast (or reduced contrast sensitivity). It is likely that many people with low vision have a visual span much less than nine characters, and sometimes as small as one character (Bullimore & Bailey, 1995; Legge et al., 1995). When Mr. Chips's visual span decreases from 9 to 1, his mean saccade size decreases by a factor of about 5 for either the noisy or noise-free cases.

Saccade Patterns: Emergent Strategies

Many studies of human reading have documented the landing sites of foveal fixations in words (cf. McConkie, Kerr, Reddix, & Zola, 1988). Mr. Chips accords no special status to the central slot of his retina. For convenience, we can label the central slot his *fovea*, and ask where it lands in words.

The top panel of Figure 5 shows results for a normal retina with a visual span of five letters. For each of three word lengths (4, 7, and 9 letters), the curves show the percentage of foveal landings on the first letter, second letter, etc. All three curves peak on the third letter, that is, Mr. Chips's fovea favors the third letter. Borrowing a term from human reading, the third letter is the *optimal viewing position* for Mr. Chips.

For human readers, the optimal viewing position is defined as the fixation position within words that yields fastest recognition (O'Regan, 1990). It lies on or just left of the center of words. Humans tend to place their fixations in a *preferred viewing position*, slightly to the left of the optimal viewing position. Of course, preferred and optimal viewing positions coincide for Mr. Chips.

For many common words (4–6 letters), the optimal viewing position would usually occur on the second or third letter. One possible reason for fixating on a specific letter position is that it may be more informative than the others.

The lower panel of Figure 5 addresses this hypothesis for Mr. Chips. For this simulation, his retina had a foveal scotoma. If the third letter of words is decisive in interpretation, Mr. Chips would rarely foveate it. Nevertheless, the curves still peak in the third letter position.

There is a different way of describing Mr. Chips's behavior. Rather than foveating a critical letter, he is left-justifying new

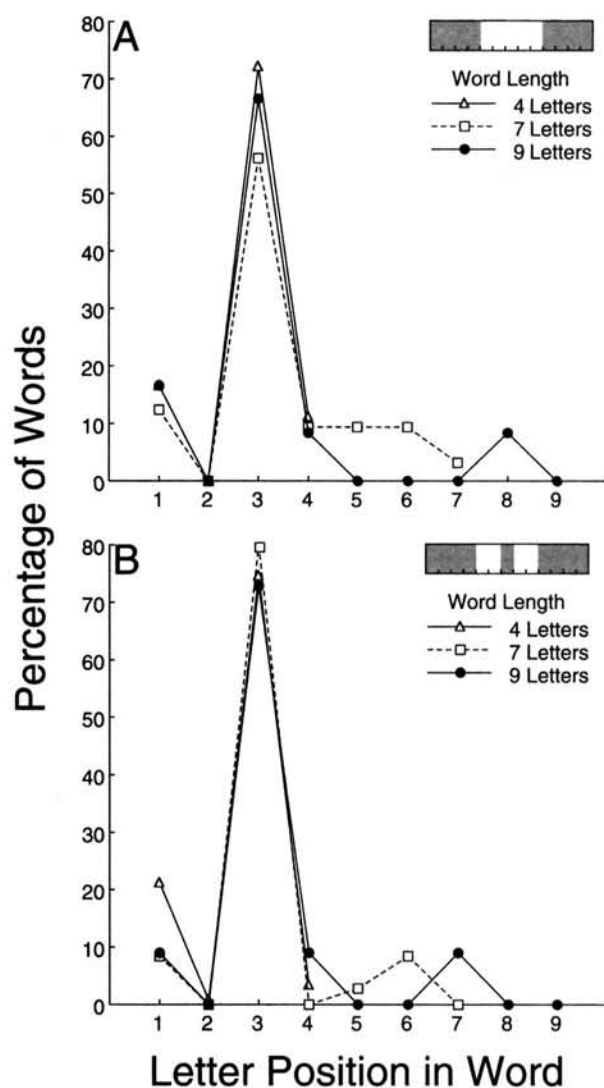


Figure 5. Illustration of Mr. Chips's optimal viewing position. Each curve shows the percentage of foveal landings as a function of the letter position within words. Separate curves are shown for word lengths of 4, 7, and 9 letters. Mr. Chips's fovea is the central slot in his retina. In A, results for a normal retina with a visual span of 5. In B, results for a retina with a central scotoma.

words in high-resolution central vision. Because the fovea is the third high-resolution slot from the left in this example, it tends to fall on the third letter. For a retina with a visual span of 7 (with or without a foveal scotoma), Mr. Chips's optimal viewing position was 4.

Although Mr. Chips's saccade decisions are governed by entropy minimization, we can summarize his behavior by saying that he often uses a strategy of left-justifying words in his visual

⁸ Angular character sizes were not given in the article, but we inferred them from stated values of saccade sizes in degrees and number of characters.

span. Humans may learn a heuristic strategy through experience: gain the maximum information about words by left-justifying them in high-resolution central vision.

We illustrate another emergent saccade strategy, this time for a fragmented retina. In Figure 6, Mr. Chips's central vision consists of four islands of high-resolution vision separated by scotomas. We term it the *Alcatraz retina*. It crudely resembles the mottling of the visual field that can result from some forms of retinal disease such as diabetic retinopathy.

Fixation 1. The first letter of *and* is in the rightmost slot of the retina. Mr. Chips makes a long saccade of 9 letters.

Fixation 2. Mr. Chips now knows that the first word has the form "a*d." The word is still ambiguous because both *and* and *add* are in the lexicon. All he can do is make a saccade of one letter.

Fixation 3. This fixation reveals the *n* of *and* and also fills in the missing letters of the next word, *could*. Mr. Chips is now ready to make another long saccade of length 9.

Fixation 4. Once again, he ends up in a situation requiring a saccade of one.

This unusual pattern of saccades—long, short, long, short—emerges from Mr. Chips's entropy computations. It is possible that humans with an Alcatraz retina could learn to use an alternating long–short strategy, without going through explicit entropy calculations.

Mr. Chips's Fovea and Parafovea

There is a prevailing view that the fovea and parafovea play unique and critical roles in reading. The fovea is believed to

recognize the word of immediate interest, with the letter at the point of fixation having special importance. The parafovea collects preview information about the word immediately to the right of the fixated word. There is evidence that this parafoveal preview facilitates recognition of the next word once it is fixated (Rayner, 1978).

There are three problems with this conventional view. First, it implies unique roles for the anatomically defined fovea and parafovea. In fact, many aspects of reading are invariant over a range of character sizes. For large letters, say 2° in width, a fixated word may cover both the fovea and parafovea, and the adjacent word may lie entirely outside the parafovea.

Second, although we can define the point of fixation as the target location imaged at the center of the fovea, it is not clear that this target point has any unique information-processing status. In fact, the idea of visual span implies the existence of a region surrounding the point of fixation in which equivalent information processing occurs. The point of fixation can be used as an arbitrary spatial reference for tagging fixations without implying special processing at that point.

Third, the notion of parafoveal preview confounds the roles of adjacent retinal regions with the processing of adjacent words within text. For Mr. Chips, and likely for humans, the properties of preview depend on the extent to which the word next to the fixated word encroaches into the visual span.

The word under scrutiny by Mr. Chips, termed the current word in the description of the algorithm, often, but not always, contains the foveal slot. For example, during the second fixation by the Alcatraz retina (Figure 6), information processing began

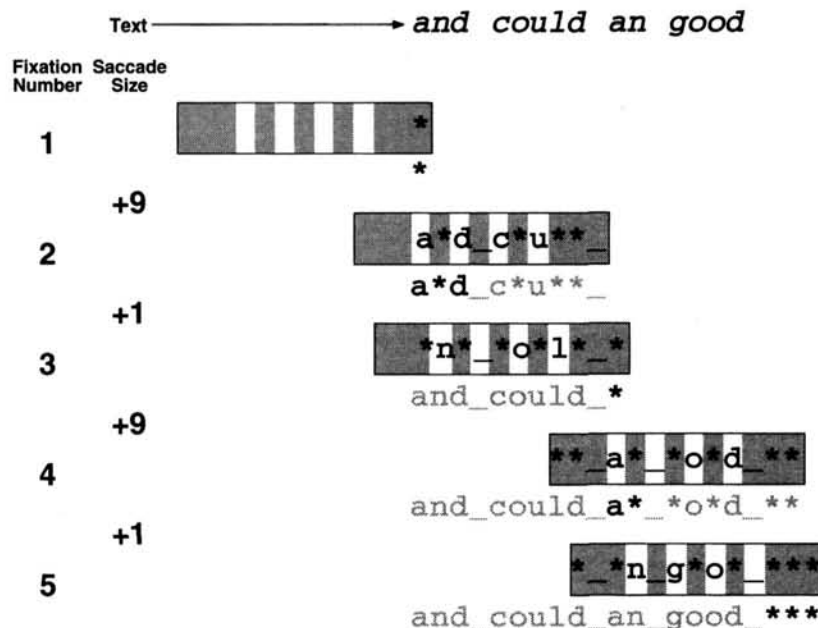


Figure 6. Series of five fixations by Mr. Chips. He has a fragmented or Alcatraz retina, in which islands of high-resolution vision alternate with scotomas. Notice that Mr. Chips adopts an unusual pattern of alternating short and long saccades. A sample animation for Figure 6, demonstrating a sequence of Mr. Chips's saccades, is available at the following WorldWideWeb address: <http://vision.psych.umn.edu/chips.html>. * = letters have been detected but not recognized; _ = a space has been recognized.

on the word *and* to the left of the fovea, proceeded to the word *could*, which contained the fovea, and terminated with consideration of the word following *could*. For Mr. Chips, the idea of sampling or "uncovering" the current word with a retinal window replaces the idea of impaling the word with the fixation point.

The human literature treats the fixated word as a central theoretical construct. According to the standard view, the fixated word receives most of the information-processing resources (cf. Just & Carpenter, 1980). Presumably, skipped words receive relatively little attention. Figure 7 shows the percentage of words skipped (i.e., not fixated) as a function of word length. The human data are replotted from Rayner and McConkie (1976, Table 1) and come from 10 undergraduates who read simple passages for comprehension. Mr. Chips's data are for a visual span of 9. A skipped word is one for which the central slot in his retina (i.e., his fovea) never comes in contact with the word. There is remarkably good agreement between the human and ideal data for word lengths up to 7. For longer word lengths, Mr. Chips skips a higher proportion of words than humans. The difference may be due to the small number of long words in Mr. Chips's lexicon and his ability to use lexical inference. The important point is that although Mr. Chips fails to fixate many words (he skips them), he devotes equal computational resources to all words. From his perspective, there is nothing special about the foveated words.

Parafoveal preview means that some information about the word next to the fixated word can be absorbed prior to a saccade

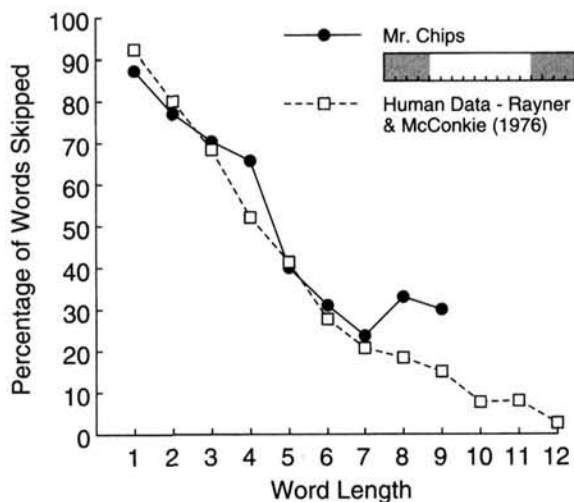


Figure 7. Percentage of words skipped (not fixated) as a function of word length. The human data are replotted from Rayner and McConkie (1976). Mr. Chips's data were obtained for a retina with a visual span of 9. A skipped word is one for which the central slot in his retina never comes in contact with the word, even though he identifies the word. Each data point for Mr. Chips is based on at least 79 occurrences. Percentages for word lengths from 1 to 7 letters were based on a text of 2,000 words. Additional simulations were required to obtain enough occurrences of 8- and 9-letter words. Although Mr. Chips's lexicon contains five 10-letter words, the low frequency of occurrence made it impractical to include 10-letter data.

and then used in recognition following the saccade. In a study by Rayner, Well, Pollatsek, and Bertera (1982), text outside of a "moving window" was mutilated such that the fixated word was preserved plus part or all (or none) of the letters in the next word to the right. They found that the leading two or three letters of the word adjacent to the fixated word influenced reading speed. Letters beyond the third had little effect.

Mr. Chips shows an analogous preview effect. Depending on the size of the visual span, one or more of the leading letters of the word adjacent to the fixated word may be identifiable. For example, suppose Mr. Chips's visual span is seven letters wide. From the optimal-viewing position principle, he tends to fixate on the fourth letter of words. If the fixated word has four letters, a space and the leading two letters of the next word lie in his visual span.

There is one way in which Mr. Chips may have an advantage over humans in the use of preview information. He readily uses information about the first, second, or any available letter or letters in the previewed word. According to Rayner, McConkie, and Zola (1980), human preview is only helpful when the leading letter of the previewed word is encoded.

There are two ways in which preview information may be useful. First, Mr. Chips uses preview information—any known letters and word length—to plan the saccade that minimizes expected entropy about the current word. For example, if word suffixes resolve uncertainty, he plans a saccade to reveal the suffix of the word. There is disagreement in the human literature about whether infrequent letter combinations at the beginning or end of preview words influence human eye movements (Hyona, 1995; Rayner & Morris, 1992; Underwood, Clews, & Everatt, 1990).

Second, in human reading, previewed words require less fixation time (Rayner, 1978). Mr. Chips also exhibits reduced processing time because the letters seen in preview greatly reduce the number of candidate words associated with the current word. But, for reasons described in the Method: General section, a theory of fixation time is not part of the computational theory of Mr. Chips. The idea of using preview to winnow down the size of the cohort of alternatives is similar to the information accrual hypothesis described by Lima (1993).

Mr. Chips encompasses the notions of foveal fixation and parafoveal preview within the single concept of visual span. This more general framework emphasizes the parallel-processing capacities of spatial vision. One advantage of this framework is that the analysis generalizes easily to the case of a retina with scotomas. No additional assumptions are needed to deal with fixation and parafoveal preview.

Retinas With Central Scotomas

Age-related macular degeneration (AMD) is the leading cause of low vision and often results in central scotomas (Ferris, 1983). People with central loss usually have severe reading deficits (Bullimore & Bailey, 1995; Faye, 1984; Legge, Rubin, et al., 1985; Whittaker & Lovie-Kitchin, 1993).

Patients with central scotomas usually adopt a nonfoveal retinal region for fixation, termed the *preferred retinal locus* or PRL (Cummings, Whittaker, Watson, & Budd, 1985; Timberlake et

al., 1986). Patients appear to adopt a PRL quickly and spontaneously (although some adopt multiple PRLs). A study of bilateral foveal lesions in monkeys indicated that they spontaneously adopted stable PRLs within 1 day (Heinen & Skavenski, 1992). For a thorough review of the PRL, see Schuchard and Fletcher (1994).

Some low-level visual capacities, such as grating contrast sensitivity, can be equated in central and peripheral vision if stimuli are scaled in size (magnified) according to the cortical-magnification factor or some other scaling rule (Rovamo & Virsu, 1979; Virsu & Rovamo, 1979). If peripheral vision was simply a scaled version of central vision, we could remedy reading difficulties in patients with central scotomas by presenting suitably enlarged text to the PRL. Magnification does help, but alas, it falls far short of restoring patients to normal reading speed.

Five factors may help explain poor reading in patients with central scotomas:

1. *Concomitant disease.* In some cases, there may be concomitant damage to peripheral retina.

2. *Location of the PRL.* The PRL is usually adjacent to the scotoma. A recent report indicated that the PRL is often located to the left of the scotoma in the visual field, seemingly maladaptive for reading (Guez, le Gargasson, Rigaudiere, & O'Regan, 1993), but there is more discussion on this later. Other authors have not observed consistent placement of the PRL relative to the central scotoma. Timberlake, Peli, Essock, and Augliere (1987) showed that the reading speeds of two patients improved when they were instructed to use a more appropriate PRL.

3. *Maladaptive saccades.* It appears that the saccade system has difficulty in adapting to using the PRL as a reference (Peli, 1986; White & Bedell, 1990; Whittaker, Cummings, & Swieson, 1991). Heinen and Skavenski's (1992) monkeys had maladaptive saccades for a long period following foveal ablation, and some never adapted successfully. White and Bedell (1990) showed that only about one third of their patients with central scotomas used the PRL as an oculomotor reference, and most of those who did had juvenile forms of macular degeneration.

Maladaptive saccades cannot be the whole story. Even when text is presented using the RSVP method to obviate the need for eye movements, reading is slower in the periphery (Rubin & Turano, 1994).

4. *Reduced visual span.* Because visual acuity peaks in the fovea and decreases in all directions, the maximum visual span should occur for a string of letters centered on the fovea. The site of the PRL relative to the scotoma may also determine the size of the visual span.

5. *Inferior pattern analysis.* There is evidence for inferior coding of pattern information in the periphery, including increased positional uncertainty (Hess & Hayes, 1994; Klein & Levi, 1987), coarser coding of spatial phase (Bennett & Banks, 1991), and difficulty with simple pattern matching (Schlingensiepen, Campbell, Legge, & Walker, 1986).

Understanding the reading deficits of people with central scotomas is a key health problem. Although a full understanding of these five factors goes well beyond the scope of Mr. Chips, the modeling in this paper can help to provide a theoretical framework for future research.

Figure 8 shows Mr. Chips's saccade histograms for four retinas with different patterns of central scotomas. The presence of a one-letter scotoma (Figure 8A) has a relatively weak effect on Mr. Chips's mean saccade length, reducing it by only 11% compared with a corresponding retina with intact central field (Figure 4C). But, as central vision becomes more fragmented, the saccades tend to be shorter and more variable in size, and there are more regressions.

There have been very few studies of reading eye movements in people with central scotomas because it is difficult to calibrate eye trackers without stable central fixation. Bullimore and Bailey (1995) measured the eye movements of 13 patients with AMD and six age-matched normal participants. Stimuli consisted of continuous text or unrelated words. The authors calculated reading speed as the product of three components: saccade frequency (number of saccades per second), ratio of the number of forward saccades to the total number of saccades, and letters per forward saccade (closely related to mean saccade length). Using regression analysis, they asked which of these factors accounted for reading deficits in patients with AMD. Most of the variance was accounted for by letters per forward saccade. For continuous text, normal participants had a mean of 7.5 letters per saccade. Most of the patients with AMD had means of 1 to 4. Trauzettel-Klosinski, Teschner, Tornow, and Zrenner (1994) measured eye movements in reading for 10 normal participants and 11 participants with juvenile macular degeneration (JMD). The pertinent means for normal participants and participants with JMD were, respectively: reading speed 269 and 112 words per minute; saccade size 9.0 and 5.1 letters; and proportion of regressions 4% and 15%. Both of these studies show that a reduction in mean saccade length plays a pivotal role in reading for patients with central scotomas.

Several authors have commented that the patterns of saccades in patients with central scotomas often appear erratic or irregular, with more regressions and a broader distribution of saccades than normal. Is it appropriate to train such patients to use more normal eye-movement patterns in reading? When Mr. Chips is obliged to use a fragmented retina, his pattern of saccades is erratic in the same sense, with more regressions and short saccades (cf. Figures 8C and 8D). It is possible that the erratic behavior of the patients is not maladaptive but has developed as a good eye-movement strategy, linked to the spatial distribution of scotomas.

Mr. Chips's behavior with central scotomas raises two questions regarding human behavior.

Do humans preserve a perceptual representation of space across scotomas? Suppose that the word *cataract* is seen in the presence of a five-letter-wide scotoma. A space-preserving representation of the stimulus would be "ca_____t" in which the underscores represent a spatial extent. If perceptual space is sewn shut across the scotoma, the word representation would be *cat*. Clearly, retention of the spatial information, even in the absence of letter information, would be helpful in reading.

Ramachandran (1992, 1993) has provided compelling demonstrations for the existence of a low-level sensory representation of the blind spot in normal vision. The blind spot is a naturally occurring scotoma in peripheral vision. He has also reported some perceptual "filling in" of scotomas in patients

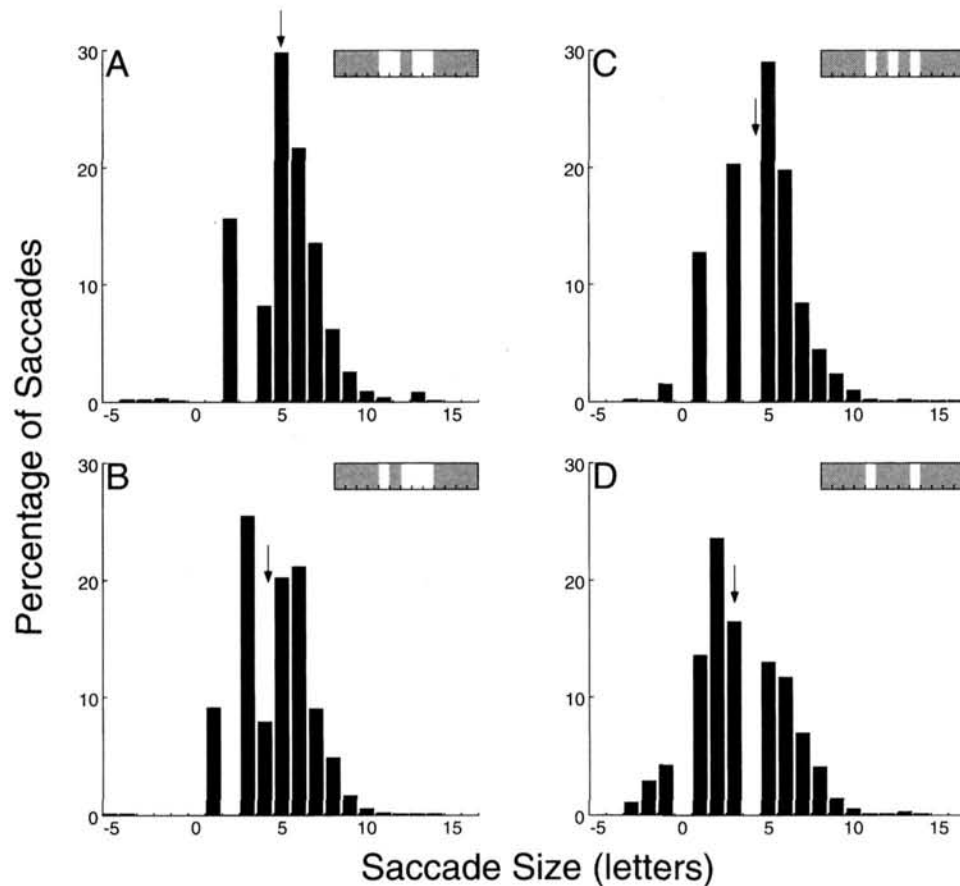


Figure 8. Saccade histograms for four retinas with different patterns of central scotomas. Each distribution was based on a text of 2,000 words.

with cortical lesions. Visual neglect, due to stroke, may be an instance in which visual space is not preserved in the perceptual representation of words (cf. Caramazza & Hillis, 1990).

Are humans able to integrate letter information across large scotomas? Consider the widely separated islands of vision in the retina of Figure 8D. Mr. Chips happily samples the text, recognizing two letters at a time, separated by a three-letter scotoma. It is unknown whether humans can make effective use of this kind of dispersed sampling. There are two key questions: How much is gained by using information from the islands on both sides of the scotoma rather than just one? If information can be used on one side only, which side yields better performance?

To obtain Mr. Chips's answers to these questions, we compared saccade histograms for retinas with a single island of high-resolution vision on the left, a single island on the right, islands on both the left and the right, and a normal retina. We did so for islands that were either one or two letters wide. Table 3 summarizes the findings. It is clear that Mr. Chips does much better when he uses both islands (Left + Right) rather than one (~ 70% increase in mean saccade size). There is an interesting difference in his behavior for a single left versus a single right island. Although the mean saccade lengths are similar, the variability and proportion of regressions are both much greater when

the island of good vision is to the right of the scotoma. When the high-resolution island is to the left of the scotoma, Mr. Chips uses both the scotoma region and the right periphery to collect word-length information. In effect, Mr. Chips has an enlarged right peripheral field.








This difference between left and right islands may explain Guez et al.'s (1993) puzzling finding that many patients place their PRL to the left of scotomas. If the scotoma retains some coarse pattern resolution, enough to tell a space from a letter, the scotoma may function in finding word boundaries to the right of the PRL.

Retinas With Peripheral Scotomas

Diseases such as glaucoma and retinitis pigmentosa can produce substantial loss of peripheral vision before they have much effect on central vision. In advanced cases, all of the visual field is affected, apart from a small island or "tunnel" at the center. Eventually, there may be a reduction of acuity within the tunnel. In these advanced cases, reading is disrupted. In some cases of stroke, the entire left or right side of the visual field can be lost (hemianopia).

Figure 9 shows saccade histograms illustrating how the loss

Table 3
Saccade Distribution Parameters for Retinas With Islands

Island	Retinal structure	Mean length	SD	% regressions
Left		2.09	1.59	8.93
Right		1.93	3.18	35.0
Left + Right		3.40	2.75	8.14
Left		3.24	1.39	0.55
Right		3.23	2.01	5.32
Left + Right		5.24	2.10	0.66
Normal		5.97	1.36	0.27

of peripheral information affects the behavior of Mr. Chips. The distribution for a retina without peripheral zones (Figure 9B) is narrower than the corresponding normal retina (Figure 9A). Some longer saccades are present in Figure 9B related to word-length information available from the right periphery of the normal retina. Notice that the difference in mean saccade lengths is quite small—5.975 in Figure 9A and 5.426 in Figure 9B.

Figures 9C and 9D compare right and left hemianopias. When

the right field is lost, reduced word-length information results in shorter and more variable saccades with a substantial number of regressions. This saccade distribution also has a few extra-long rightward saccades. Often a regression is followed by an extra-long rightward saccade: Mr. Chips remembers the intervening characters and jumps over them. (There is evidence from human readers with normal vision for long rightward saccades following regressions, Rayner, 1993.)

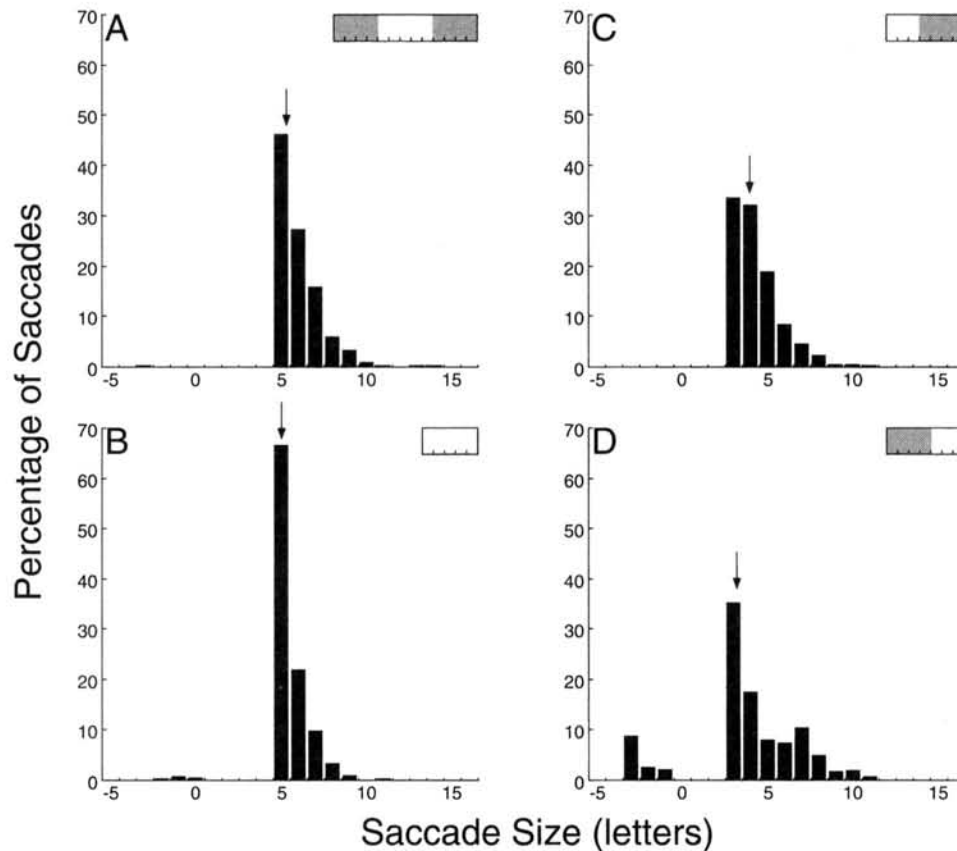


Figure 9. Saccade histograms for a normal retina (A), a retina with no peripheral vision (B), a left hemianopia (C), and a right hemianopia (D). Each distribution was based on a text of 2,000 words.

The right hemianopia has another effect on Mr. Chips that is not evident from the histogram. The simulation's computing time grows longer. Mr. Chips uses word-length information (based on detection of bounding spaces in the right periphery) to reduce the number of candidate words by a large amount (see the Method: Mr. Chips's Algorithm section). It is clear that an ideal reader that relies on serial computation would benefit in speed from access to spacing information.

In the Discussion section, we consider the role of the left peripheral field in fixation linking.

Text Segmentation

Mr. Chips uses two restrictions to simplify segmentation of the text into words: the guarantee of a single space between words and inclusion of relative, but not absolute, scotomas.

Spacing. For Mr. Chips, a space is a reliable segmentation marker and guarantees that a unique interpretation of each word is possible. If spaces were removed, Mr. Chips would have to rely on knowledge of the set of possible words for segmentation. Unresolvable ambiguities could occur. For example, he could never be sure whether *become* was a single, six-letter word or the two words *be come*. Of course, Mr. Chips could use his knowledge of the word frequencies to make a best guess at the correct segmentation. In principle, it would be straightforward to design a variant of Mr. Chips to read unspaced text.

Humans can read unspaced text. Most segmentation ambiguities can be resolved from context, although it is possible to contrive ambiguous strings such as *AMANDASHEDHEREARN-ESTILLGOTTENANTS* or *IFINDIANSWEREDONE* (courtesy of Andrew Luebker). There is evidence that humans slow down when they read unspaced text. Estimates of the mean speed-reduction range from about 30% to 50% (Epelboim, Booth, & Steinman, 1994; Spragins, Lefton, & Fisher, 1976). It is possible that the additional cognitive demand required for segmenting unspaced text results in slower reading or reduced comprehension.

Absolute scotomas and the multiple-segmentation problem. The software implementation of Mr. Chips permits relative scotomas anywhere in the retina: slots in which a space can be discriminated from a letter, but letters cannot be recognized. One step in generalizing Mr. Chips would be to permit absolute scotomas: slots with no vision whatsoever. Absolute scotomas introduce an extra level of computational complexity.

Suppose Mr. Chips's retina contains two absolute scotomas, in addition to relative scotomas. Figure 10 shows the ambiguity associated with multiple segmentation. Each slot containing an absolute scotoma might hide a space. With two absolute scotomas, there are four possible segmentations of the text. In the general case of N absolute scotomas, there are 2^N possible segmentations.

How would an ideal reader handle multiple segmentation? Often, some of the branches could be pruned using lexical constraints or spacing constraints (i.e., two spaces cannot occur side by side). But sometimes, Mr. Chips would face two or more plausible segmentations. In Figure 10, he knows that the current word is either *car*, or a seven-letter word of the form "car*fu*" or a nine-letter word of the form "car*fu***."

Actual text:	<i>carefully_</i>
<u>Absolute scotomas & relative scotomas</u>	
Visual output	<i>car•fu***_</i>
Four possible segmentations with examples:	<i>xxx_xxx_x_</i> <i>car_fun_a_</i>
	<i>xxx_xxxxx_</i> <i>car_fully_</i>
	<i>xxxxxxxx_x_</i> <i>careful_a_</i>
	<i>xxxxxxxxxxx_</i> <i>carefully_</i>
<u>Relative scotomas only</u>	
Visual output	<i>car*fu***_</i>
One possible segmentation:	<i>xxxxxxxxxxx_</i> <i>carefully_</i>

Figure 10. Illustration of the multiple-segmentation problem. In the examples, • codes for an absolute scotoma, _ indicates a space, and * shows an unidentified letter. The pattern of xs shows the possible segmentations of the text into words and spaces. When there are two absolute scotomas, there are four possible segmentations. When there are no absolute scotomas, there is only one possible segmentation.

From his lexicon, he can establish the sets of candidate words that fit these constraints. He would then compute the expected entropy associated with each possible saccade, averaging across segmentations. To calculate the segmentation probabilities, Mr. Chips would have to analyze properties of possible words lying to the right of the current word.

How would a human reader deal with multiple segmentations resulting from absolute scotomas? The simplest strategy is to ignore all information to the right of the first possible segmentation point (i.e., to the right of the leftmost border of any scotoma). A riskier strategy is to guess that the scotoma does not hide a space and that the correct segmentation is determined by the visible spaces. A priori, the probability of a space occurring in any particular text location is related to mean word length. Finally, it is possible that humans with scotomas have some capacity to process information about two or more segmentations simultaneously.

The multiple-segmentation problem makes reading with absolute scotomas qualitatively more complex than unimpaired reading.

Results: Lexical Analysis

Figure 2 contained an example in which Mr. Chips saw "diffe****." The partial information was sufficient to elimi-

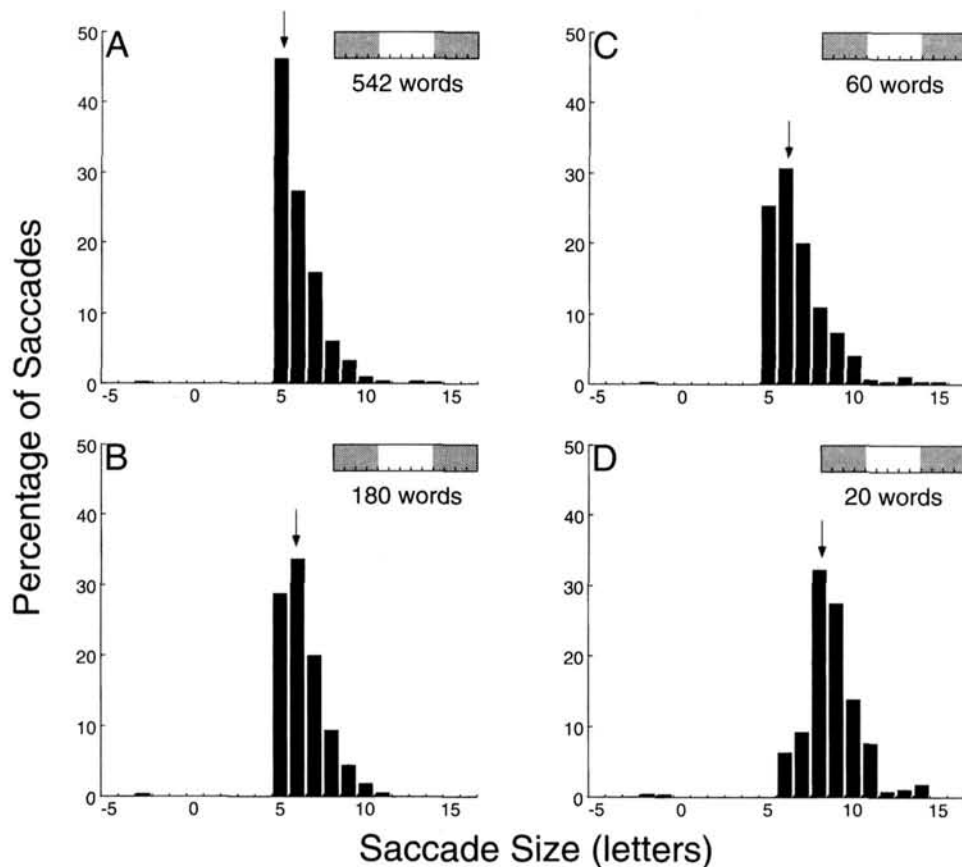


Figure 11. Saccade histograms for lexicons with 542, 180, 60, and 20 words. Each distribution was based on a text of 2,000 words.

nate all words in his 542-word lexicon except *different*. Lexical inference enabled Mr. Chips to leap over the unseen end of the word and begin analysis of the next word.

Lexical knowledge helps Mr. Chips in two ways: in reducing uncertainty about the word under study and in planning an optimal saccade. How much advantage does he gain from his lexical analysis? The answer depends on lexicon size and the retinal sampling structure.

Lexicon Size and Context

Mr. Chips does not use context. Context, in the form of semantic and syntactic constraints, reduces the number of likely words at any point in meaningful text. In this sense, variations in lexicon size can simulate context effects. Figure 11 shows saccade histograms for four lexicons. The greater intrinsic uncertainty associated with the larger lexicons results in a reduction in mean saccade length. The uncertainties (in bits) associated with these four lexicons are 7.28 (542 words), 6.04 (180 words), 4.89 (60 words), and 3.58 (20 words).⁹

Seemingly, Mr. Chips is a victim of a vocabulary paradox in which his reading speed slows down as his lexicon size increases. There is no evidence that erudite humans with large

vocabularies read more slowly than people with small vocabularies. The paradox is resolved by understanding that it is the predictability of the text that determines mean saccade size, not the raw number of words in the lexicon. Highly predictable text could be constructed from a large lexicon: A few words would have high probabilities, and the rest would have low probabilities.

Similarly, less predictable text can be constructed from smaller lexicons. For instance, real-world texts on technical topics often use a smaller lexicon than literary prose. Nevertheless, people usually read technical prose more slowly because the semantic complexity makes the text less predictable.

Figure 12 shows how Mr. Chips's mean saccade length depends on visual span for 542- and 20-word lexicons. Comparison of Figures 12A and 12B shows that for normal retinas, the decreased uncertainty associated with the 20-word lexicon

⁹ Because the words have unequal probabilities of occurrence, the entropies are less than values for a uniform distribution. For example, the entropy associated with 542 equally probable words would be 9.08 bits, compared with the value of 7.28. Therefore, Mr. Chips gains some benefit in entropy reduction from his knowledge of word frequencies.

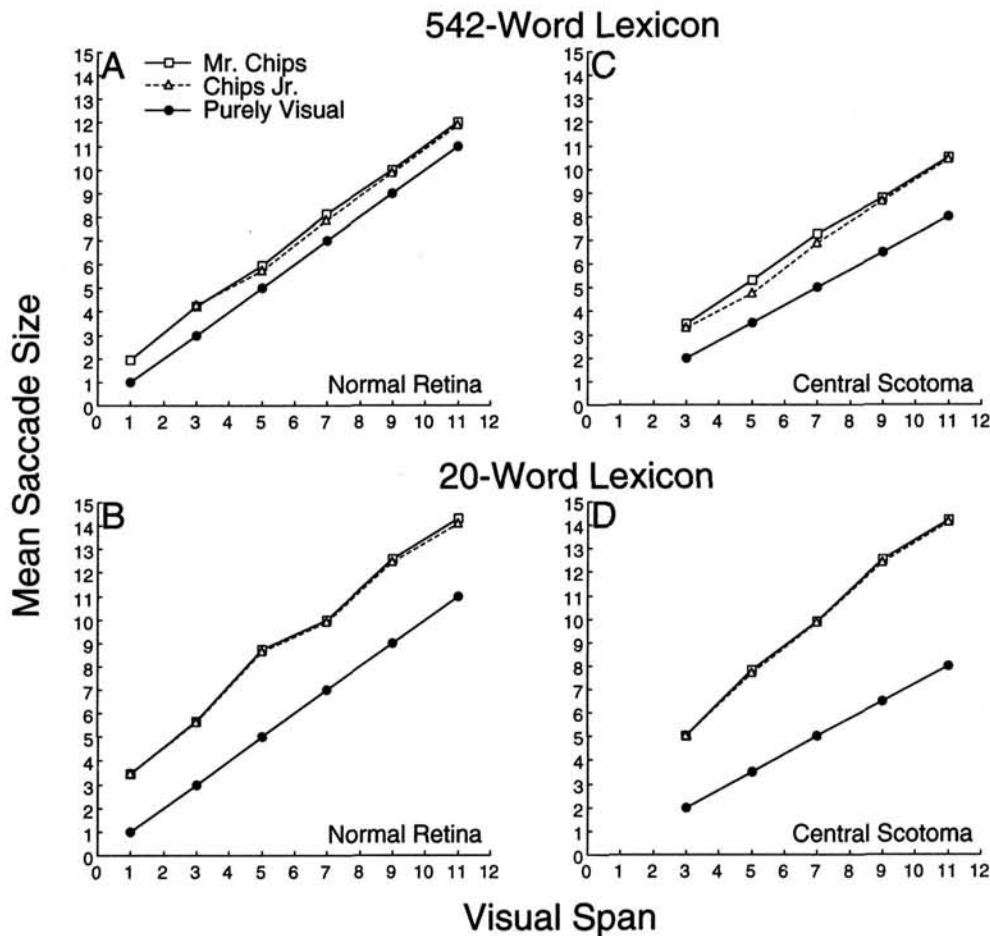


Figure 12. Mean saccade length as a function of visual span for 542- and 20-word lexicons. Figure 12A and 12B show results for normal retinas (no scotomas), and Figures 12C and 12D show results for retinas with a one-letter central scotoma. The three curves in each panel are for Mr. Chips and two suboptimal models.

translates into an increase in mean saccade length of about two characters, independent of the visual span.

Do humans make use of the predictability of text to increase reading speed? Morton (1964) measured the fixation times and the number of saccades of participants as a function of the predictability of text. The texts were constructed as zero- to eighth-order approximations to English (Miller & Selfridge, 1950; Shannon, 1951). (Mr. Chips reads first-order approximations.) As the order increases, the added contextual constraints reduce the set of likely words. Increasing the contextual constraint is equivalent to reducing the size of the effective lexicon at any point in the text with a corresponding reduction in uncertainty.

Morton's (1964) participants were told to "read the passages as quickly as possible, minimizing errors," close to the "instructions" given to Mr. Chips. Morton found that human reading speed increased by about 33% from the first-order text to the eighth-order text and that there was no difference between eighth-order text and real text. The mean fixation time was

constant, so the change was due to an increase in mean saccade length. Similarly, for a visual span of 9, Mr. Chips's mean saccade length is 25% greater for the 20-word lexicon than the 542-word lexicon (Figure 12A and 12B).

The similarity in behavior of Mr. Chips and Morton's (1964) human participants suggests that contextual effects on reading speed are fairly small and can be explained by the reduction in uncertainty associated with the moment-by-moment activation of a very small lexicon.

Humans are usually able to contend with unknown words in text. They use context or lexical knowledge to do error correction (for typographical errors or "typos") or guess the meaning of the unknown word. If context is unavailable to support these strategies (e.g., unfamiliar jargon terms), people may be left in a state of unresolved ambiguity.

How would Mr. Chips deal with an unknown word in text, that is, a string not listed in his lexicon? An extension to the model would be necessary to provide rules governing the set of possible nonwords. For instance, a typo model might permit any

letter to be replaced with other letters according to a probability distribution, and Mr. Chips might be given permission to skip a word when a typo is detected. The set of nonwords produced by typos would include a very large set of low probability strings for Mr. Chips to remember. Once he was sure a given string was in the extended set of nonwords, he could safely skip past it without completely identifying it. Other forms of nonlexical intrusions would require derivation of corresponding optimal error detection (and possibly correction) algorithms.

Reading Strategies and the Advantage of Lexical Inference

Mr. Chips's lexical inference is computationally intensive. What advantage does he gain from it?

Figure 12 compares Mr. Chips's performance with two sub-ideal cousins. The purely visual observer makes no use of lexical inference. For a normal retina with a visual span of N , repeated saccades of length N uncovers the entire text string (assuming no saccade noise). This strategy places a lower bound on the average saccade length we would expect a reader to adopt. This strategy is computationally trivial: Each saccade is the same length.

Intermediate between the ideal strategy of Mr. Chips and the purely visual strategy is a modified lexical strategy, termed *Chips Jr.* This strategy uses word-length information but no letter information. For example, if 's**_' is visible, Mr. Chips would base his analysis on the 10 three-letter words in his (standard) lexicon beginning with *s*. Chips Jr., however, uses only the information that the word has three letters. Because there are 79 three-letter words in the lexicon, it would seem that Chips Jr. would have more computation than Mr. Chips, not less. As described in the Method: Mr. Chips's Algorithm section, however, Mr. Chips uses a look-up table to store information on optimal saccades when only word-length information is available. Chips Jr. bases his saccade planning exclusively on this look-up table and does not do any entropy calculations using partial letter information.

Comparison of the performance of Mr. Chips, Chips Jr., and the purely visual observer in Figure 12A yields two noteworthy results. For the standard 542-word lexicon, all of Mr. Chips's lexical analysis only bags him one extra letter in mean saccade length over a purely visual strategy: $M = N + 1$ (Mr. Chips), $M = N$ (purely visual observer). Secondly, Chips Jr.'s performance is only slightly poorer than Mr. Chips's performance. The key point is that Mr. Chips, for all his number crunching, gains very little relative to a visual strategy that requires almost no computation. What little he gains can almost entirely be achieved by using word-length information alone.

This is an instance in which computationally simple, sub-ideal strategies can yield performance nearly as good as the ideal observer. The ideal observer achieves a small performance advantage at a large computational cost. Human readers may find it easier and faster to adopt a purely visual strategy. If so, they would make short saccades to reveal the ends of words like *different* in Figure 2, rather than spending time in contemplation (like Mr. Chips) in order to infer what word is present.

Chips Jr. is similar to the strategy-tactics model of eye-move-

ment control in reading (O'Regan, 1990). According to this model, saccade planning does not rely on detailed linguistic or lexical analysis but uses simple strategies that are based on coarse visual cues (especially word length). The strategy for moving from one word to the next is to aim the fovea for the center of the word.

The advantage of lexical inference increases when the visual span is small or the lexicon is small. For Mr. Chips, this reading advantage can be expressed as the percentage increase in mean saccade size relative to the case of a purely visual strategy. The bar graph in Figure 13 shows that for a normal retina with a visual span of 3, Mr. Chips has a reading advantage of about 42% over a purely visual observer (mean saccade sizes of 4.24 letters vs. 3.00 letters). The corresponding reading advantage for the 20-word lexicon is much larger (88%).

Implications for Low Vision

Lexical inference (and other forms of cognitive inference) may be more important in low-vision reading than in normal reading. Analysis of Mr. Chips's behavior indicates three reasons why this may be so. First, a purely visual strategy is more complicated and yields poorer performance in low vision. For example, suppose Mr. Chips has a retina with a visual span of 5 and a one-character central scotoma, as shown in Figure 3. What purely visual strategies guarantee success in seeing all characters of the text string? A sequence of 5-letter saccades would not work this time, failing to reveal letters lying beneath the scotoma. The best purely visual strategy is a repeating sequence of a pair of saccades, 2 letters and 5 letters in length, with a mean of 3.5 letters. Retinas with more complex patterns of scotomas could require even more elaborate purely visual strategies.

Secondly, as illustrated in Figure 12, there is a wider performance gap for the retina with a central scotoma between Mr. Chips and the purely visual observer; lexical inference confers a greater advantage.

Third, as described earlier in this article, some forms of low vision may result in small visual spans for reading. As Figure 13 shows, the reading advantage due to lexical inference is greater for small visual spans. Taking an extreme case, for a visual span of 3 with a central scotoma and a lexicon of 20 words, the reading advantage is about 150%. In other words, for these conditions, Mr. Chips reads between two and three times faster than a purely visual observer.

There is some evidence that a greater cognitive load is associated with low-vision reading. Dickinson and Rabbitt (1991) tested reading comprehension in normally sighted participants with simulated low vision (optical blur and distortion). Free recall performance was impaired, although performance on multiple-choice questions was not. In the Bullimore and Bailey (1995) study described earlier, patients with AMD read continuous text 2.12 times faster than random-word sequences, compared to a factor of 1.26 times for the controls.¹⁰ The greater difference for the patients implies a greater reliance on context.

¹⁰ These factors take into account the difference in mean word length between the text stimuli and random-word stimuli reported by Bullimore and Bailey (1995).

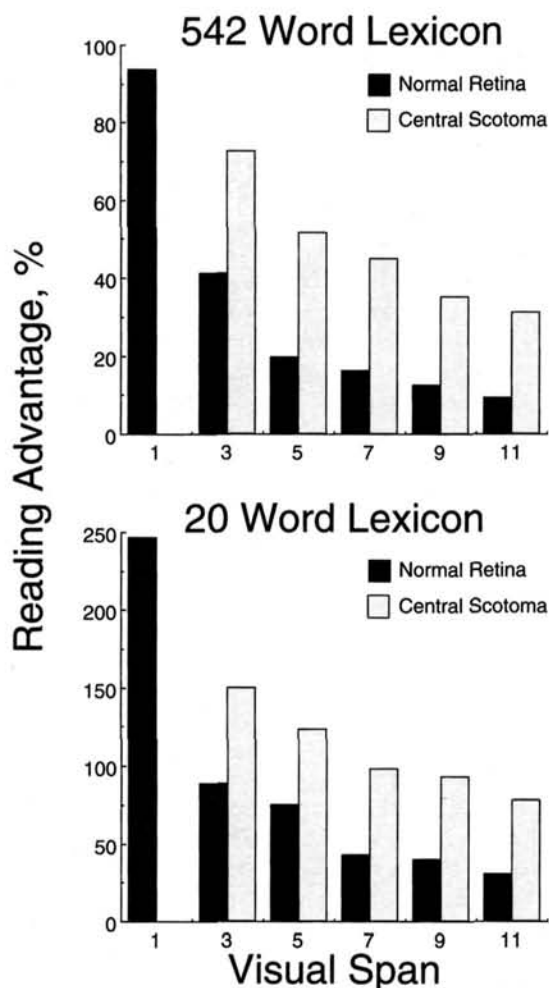


Figure 13. Mr. Chips's reading advantage in using lexical inference as a function of visual span. Values are shown for 542- and 20-word lexicons and for retinas with and without a central scotoma. No value is shown for the central scotoma when the visual span is 1 because the scotoma would obliterate the only high-resolution slot. Reading advantage is the percentage increase in mean saccade size relative to the performance of a purely visual strategy.

Cognitive load may help explain a puzzling observation from our laboratory. Akutsu, Legge, Ross, and Schuebel (1991) showed that there are only minor differences in reading speed between young and old normally sighted participants. On the other hand, age per se had a strong effect on the reading speeds of a large sample of participants with low vision (Legge et al., 1992). If lexical inference plays a more critical role in low-vision reading, then this age by low-vision interaction could be due to age-related changes in lexical access. There is evidence for deficits in word retrieval in old age (cf. Kemper, 1992).

Results: Analysis of Saccade Variability

Much debate has centered on the causes of saccade variability in human reading. Mr. Chips also exhibits variability in saccade length, but the causes are somewhat clearer.

It is important to distinguish between stochastic sources of variability (saccade noise) and deterministic sources. Retinal-sampling structure can contribute to saccade variability in a deterministic way. For instance, the Alcatraz retina in Figure 6 could uncover the entire text string in the absence of any lexical knowledge with a sequence of saccades of length 7, 1, 7, 1, etc., having a mean of 4 and a standard deviation of 3. Lexical knowledge also contributes variability.

Mr. Chips can also be asked to contend with saccade noise, analogous to oculomotor noise in humans. In the simulations, the noise consisted of a Gaussian distribution of saccade lengths (see the Method: Mr. Chip's Algorithm section). The noisiness of the saccades is specified by the standard deviation of the Gaussian. We studied two cases: *additive noise* for which the standard deviation was constant, independent of intended saccade length, and *multiplicative noise* in which the standard deviation scaled linearly with intended saccade length. The noise examples in this article use additive noise because this case is easier to understand and more informative.

Saccade Noise in Human Reading

Font noise. Mr. Chips uses number of characters as a spatial metric. In his world, each character occupies the same horizontal real estate. In the real world, this is true only for fixed-width fonts, such as Courier. In modern printing, the de facto standard for books and newspapers is a variable-width (proportional) font. There is variation in the horizontal space allotted to each letter, m taking much more than i . A fixed angular size covers a variable number of letters. A lower-bound estimate of "font noise" is the standard deviation of the letter widths in the font, expressed as a percentage of mean letter width, taking into account the probabilities of occurrence in English. For a selection of common proportional fonts (PostScript fonts in an Apple LaserWriter Plus [Apple Computer, Cupertino, CA]), the font noise ranged from 23% (New Century Schoolbook) to 34% (Avant Garde Gothic).

If saccade lengths were planned in units of letters (as Mr. Chips does), font noise would introduce some indeterminacy in the landing sites. For strings of N characters, the standard deviation of the mean string length is $N^{1/2}$ times that of an individual letter. For instance, for a font with a standard deviation of 25% for a single letter, the standard deviation for strings of 10 letters would be about 80% of a letter (i.e., $10^{1/2} \times 25\%$). For this font, a saccade of 10 letters would have an indeterminacy of nearly plus or minus 1 letter in its landing site. Shorter saccades would have less indeterminacy. Perhaps a defining property of a highly readable font is that its noise is less than the intrinsic noise of human reading saccades.

Despite font noise, normally sighted people read proportionally spaced text slightly faster ($\sim 5\%$) than fixed-width text (Arditi, Knoblauch, & Grunwald, 1990; Mansfield et al., 1996). Perhaps the speed advantage for the proportional font is due to packing more characters into the visual span. For participants with low vision, the reading-speed advantage goes to the fixed-width font (Mansfield et al., 1996).

Precision of reading saccades. How much error is there when human subjects try to make accurate saccades to isolated

targets? Van Opstal and van Gisbergen (1989) obtained saccade distributions for 0.5° targets in the periphery. The standard deviation increased linearly with saccade length with a Weber fraction of about 5%.¹¹ He and Kowler (1991) measured the accuracy of participants when they made a saccade to a single eccentric point or to a designated location within a triangle. Participants were told to stress accuracy, not speed. Latencies were between 250 and 300 ms. They found that saccade precision was typically about 13% (standard deviation). Accuracy was nearly as good for an interior location within a form as for an isolated eccentric target. This finding is pertinent to targeted saccades within words in reading.

Reading is a dynamic task requiring a rapid sequence of saccades, not a single saccade. It is likely that the saccade noise in reading is greater than for conditions meant to optimize saccade accuracy. Coeffe and O'Regan (1987) showed that latencies of at least 500 ms are required to achieve one-letter accuracy of saccades to a target letter within a string of letters. For latencies more typical of reading (250 ms), saccades were sometimes two or more letters off.

Parish and Legge (1988) used a psychophysical method to estimate the stochastic variability in eye placement during reading. Text on a video screen was randomly jittered, that is, displaced away from its nominal position on every video frame according to a two-dimensional Gaussian distribution. Reading speeds of participants with normal vision were little affected by substantial amounts of jitter, implying a considerable amount of imprecision in the placement of eyes during reading. Using an "equivalent noise" analysis, they estimated the *internal jitter*. In agreement with Coeffe and O'Regan (1987), they estimated a horizontal position error (standard deviation of landing sites) equivalent to about 2.2 letters, independent of angular character size. For saccades of average size (seven letters), indeterminacy of 2.2 letters is about 30% of the mean.

A typical estimate of total saccade variability in human reading is 40% of the mean (O'Regan, 1990). The preceding review suggests that a major contributor to this variability is inherent saccade noise, with the remainder caused by deterministic (linguistic and visual) factors.

In the Parish and Legge (1988) study, participants with low vision showed a more rapid decline in reading speed for small amounts of text jitter. The surprising implication is that participants with low vision have greater absolute precision in eye placement (less saccade noise). Parish and Legge estimated a mean stochastic position error of about 1.4 characters. Evidence cited earlier indicates that low-vision saccades are shorter than normal and lower saccade noise may be a consequence of shorter saccades.

The idea that low-vision saccades may actually be less noisy than normal saccades flies in the face of the common view that inadequate eye-movement control explains poor reading in participants with central scotomas. We are not aware of eye-movement measurements of the precision of saccades in patients with central loss.

Mr. Chips and Saccade Noise

How does Mr. Chips deal with saccade noise?

Figure 14 shows a sequence of fixations when Mr. Chips has

a visual span of 5 and a one-character central scotoma. The saccade noise has a standard deviation of 1.0. The example illustrates a key point: Although Mr. Chips plans his saccades using the principle of entropy minimization, his behavior can be summarized by a handful of heuristic strategies (Table 4). A fixation-by-fixation commentary follows.

Fixation 1. Mr. Chips wants to left-justify the unknown word in his visual span. This strategy is listed in Table 4 as the *left-justify (LJ)* strategy. He tries a saccade of eight letters. Saccade noise carries him one character further than intended.

Fixation 2. He sees "he." Because the word could be *she* or *the*, he needs to check the first letter. He attempts a regression of -2 , to place the unknown letter in the retinal slot one left of the scotoma.¹² Table 4 lists this strategy as the *single-letter (SL)* strategy. Addition of a noise sample of $+2$ to the intended regression of -2 results in a net zero saccade.

Do humans have zero-length saccades, that is, abortive saccades that go nowhere? It would be difficult to tell behaviorally.

Fixation 3. Following the zero saccade, Mr. Chips once again tries a -2 regression. This time, the noise sample of -2 results in a net regression of -4 .

Fixation 4. Although he loses ground, he does get the *t* of *the*. From Fixation 2, Mr. Chips remembers that the next word is *it* and the word following has two letters. He attempts a saccade of 9, which would place the second letter of the unknown two-letter word in the scotoma. We refer to this as the *two-letter word (2W)* strategy. It is counterintuitive that Mr. Chips deliberately aims his scotoma for the two-letter word when there are pairs of high-resolution slots on either side of the scotoma. Noise considerations once again account for this behavior.

Fixation 5. Unfortunately for Mr. Chips, the noise sample was zero, so the scotoma obscures the second letter. At this stage, Mr. Chips needs to know the second letter, so he implements the SL strategy.

Fixation 11. Mr. Chips encounters something new. His lexical knowledge tells him that "tw*" must be the word *two*. The next word begins with *h* and is three letters long. There are seven possibilities: *has, had, her, him, his, hot, and how*. The SL strategy does not apply because identification of neither the second nor third letter guarantees zero uncertainty. Mr. Chips attempts a saccade of plus three. Zero noise would leave the word straddling the central scotoma. We refer to this as the *three-letter word (3W)* strategy.

Fixation 12. Fortunately, the noise sample is -1 , so Mr. Chips sees the remaining letters of *his*.

Further study indicates that Mr. Chips also has a *four-letter word (4W)* strategy for cases such as "s***_." He targets the third letter on the scotoma.

¹¹ The distribution of saccade endpoints is actually two-dimensional, radial and tangential, corresponding to positions along a line of text or perpendicular to it. Van Opstal and van Gisbergen (1989) found that the radial SD was about twice as large as the tangential SD.

¹² Why did Mr. Chips not try to left-justify the target letter in the visual span? Consideration of the effects of the Gaussian noise shows that the odds of the critical letter landing in a high-resolution slot are slightly higher when the slot next to the scotoma is targeted.

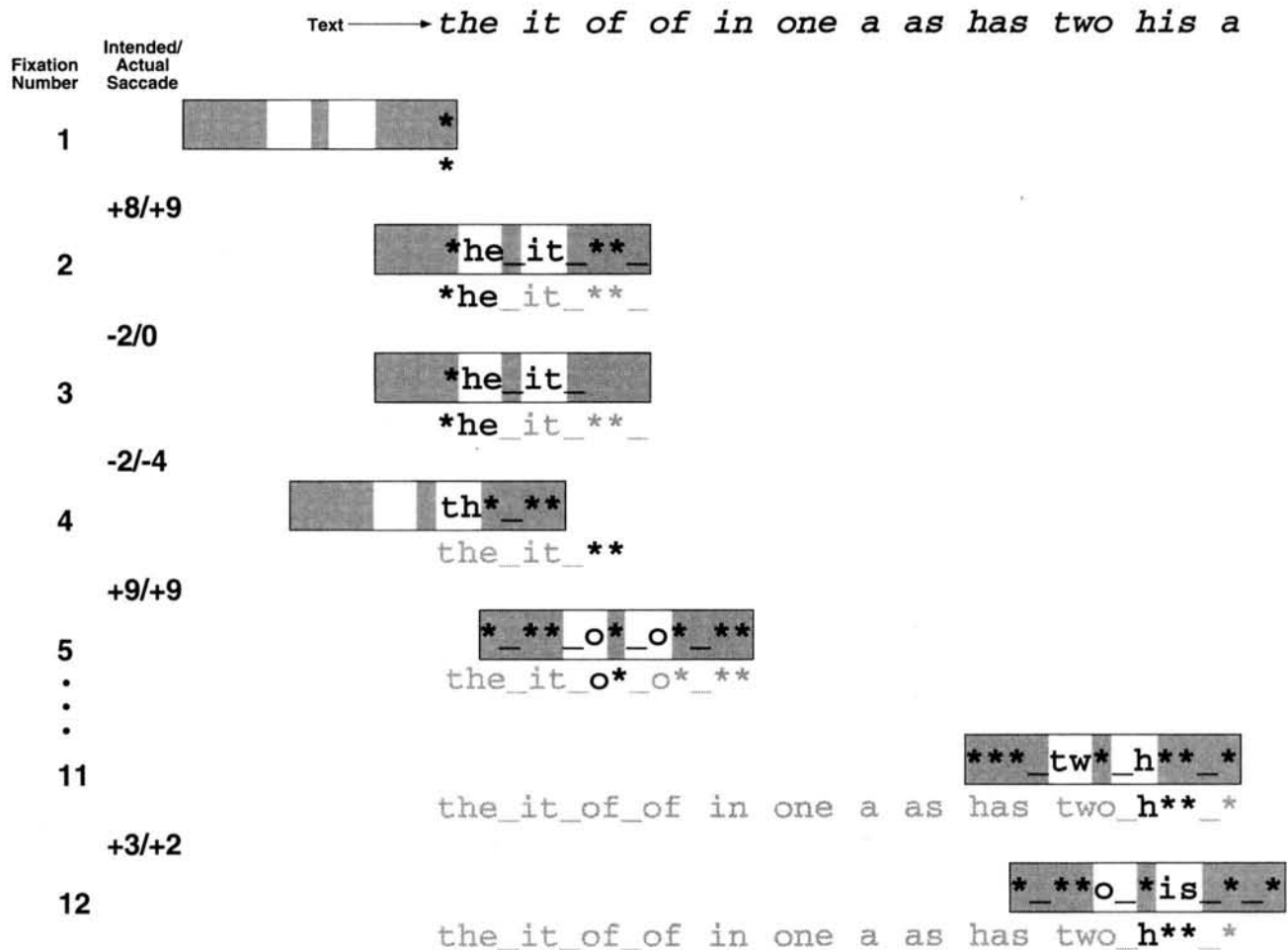


Figure 14. Series of fixations in the presence of saccade noise when Mr. Chips has a one-character central scotoma. The noise had a standard deviation of 1.0 character, independent of saccade length. As indicated in the figure, actual saccades sometimes had a different length from intended saccades. As described in the text, Mr. Chips adopted strategies to combat the effects of saccade noise. * = letters have been detected but not recognized; _ = a space has been recognized.

Three additional heuristics were added to complete the set in Table 4: single-letter modified (SLM), left-justify modified (LJM), and five-letter word (5W). These are extensions of the others and were included to cover all cases encountered in a lengthy example.

Notice from Table 4 that the 2W, 3W, 4W, and 5W strategies all tend to center words in the visual span, despite the presence of the scotoma. The four strategies comprise a *master centering strategy*. For words of known length, but missing letters, it is best to center the word in the visual span to protect it against noise. This provides another insight into the tendency of human readers to fixate near the center of words.

The appropriate set of heuristics for describing Mr. Chips's behavior would depend on the details of the retinal-sampling structure, the lexicon, and the properties of the saccade noise. A priori, we should not assume any particular set of heuristics to be universal.

A key point from the example is that Mr. Chips's behavior can be summarized by a short list of simple heuristics, which are given in Table 4. These heuristics are not an explicit part of his computational competence, but emerge from the use of the entropy-minimization principle.

A nonideal reader, facing the same constraints, could learn to use this set of heuristics. How well would such a reader perform compared to Mr. Chips? We conducted a comparison on a sample text of 100 words. The heuristic reader's mean saccade length was 3.45 letters ($SD = 3.37$) compared with Mr. Chips's mean of 3.51 letters ($SD = 3.50$). Mr. Chips took 129 saccades, and the heuristic reader 131. In short, the heuristic reader achieved nearly ideal performance without doing a single entropy calculation. The column headed N in Table 4 lists the number of occurrences of each heuristic in the sample run.

O'Regan (1990) has proposed that humans make use of a set of strategy tactics to guide their eye movements in reading,

Table 4
Mr. Chips's Heuristic Strategies

Strategy name	<i>N</i>	Description
Single letter (SL)	53	When a single letter resolves all ambiguity in a word, try to put it in the slot immediately to the left of the scotoma.
Single-letter modified (SLM)	5	If identification of a single letter in two or more positions in the word resolves all ambiguity, apply SL to the rightmost case.
Left-justify (LJ)	33	Left-justify a word of unknown length in the visual span.
Left-justify modified (LJM)	4	If leading letters are known in a word of unknown length, apply LJ to the unknown portion of the word.
Centering		
Two-letter word (2W)	12	Put the second letter of an unknown two-letter word in the scotoma.
Three-letter word (3W)	14	Put the central letter in the scotoma.
Four-letter word (4W)	8	Put the third letter in the scotoma.
Five-letter word (5W)	2	Put the central letter in the scotoma.

Note. These strategies apply when Mr. Chips has a five-letter visual span with a central one-letter scotoma. The saccade noise has a standard deviation of 1.0, independent of saccade length. *N* refers to the number of times the strategy was used in a sample run.

similar in spirit to the heuristics discussed here. Instead of taking the time to compute precise saccades based on detailed visual and linguistic analysis of words, O'Regan proposes that oculomotor control in reading uses coarse visual cues (such as word length) and simple strategies. A centering strategy aims the eyes toward the center of words, based on detection of word length. This strategy is similar to Mr. Chips's master centering strategy.

Figure 15 shows a peculiar sequence of saccades when Mr. Chips has an Alcatraz retina. The saccade noise has a constant standard deviation of 1.6.

Fixation 1. Mr. Chips is working on the text word *world*. He knows the leading letter *w* and that the word has at least four letters. He attempts a saccade of length 5. A saccade of 3 or 7 might seem to yield the same information, but these alternatives are less resistant to noise. The noise sample is -2 so the net saccade is $+3$.

Fixation 2. Mr. Chips learns two more letters and the word length. The only two consistent words in his lexicon are *world* and *would*. To resolve this ambiguity, he must identify the third letter. He tries to do this with an intended saccade of $+2$, which would place the most central scotoma directly over the critical letter. With the noise sample of $+2$, the net saccade was $+4$. Here again, Mr. Chips adopts the counterintuitive strategy of trying to place a critical letter in a scotoma. This is because his odds of identifying the key letter are slightly better when he aims for the scotoma.

Sometimes, people with macular disease bring targets of interest into the central scotoma. This is universally regarded as a maladaptive foveal reflex. Although this view is probably correct, Figure 15 (and the centering strategies in Table 4) show that an optimal strategy sometimes involves centering unresolved words in the retina, even if there is a central scotoma. This general centering strategy makes Mr. Chips less vulnerable to noise.

Fixation 3. Mr. Chips learns nothing new, and plans a regression of -2 . The noise is 0, so the intended regression of -2 is achieved.

Fixation 4. Curiously, although he hits his target dead on, the word remains ambiguous. Success is failure. The unknown third letter of the word lands in the scotoma at the center of the retina. Where should he aim for next? Taking the noise distribution into account, his retina is already in the best possible place; there is no better retinal-sampling position. Mr. Chips makes the surprising decision to attempt a zero-length saccade. He relies on the probability distribution of the additive noise to offset his retina such that the critical letter ends up on a high-resolution slot. He attempts a zero saccade, but the noise value is $+2$, for a net saccade of $+2$.

Fixation 5. Once again, the retina is in a location that provides no new information.

Mr. Chips hunts back and forth like this several more times in search of the elusive third letter. Ultimately, he makes 9 saccades to figure out *world*.

The example in Figure 15 illustrates two bizarre aspects of Mr. Chips's behavior: He attempts zero-length saccades, and he tries to place a critical letter in a scotoma. In both cases, Mr. Chips relies on the error statistics to get the letter into a high-resolution slot.

Finally, we consider how deterministic and stochastic variability combine statistically.

Figure 16 shows the effects of different levels of additive saccade noise on the mean and standard deviation of saccade distributions. Plots are shown for a normal retina with a visual span of 5 and for the Alcatraz retina. Labeled lines in Figures 16C and 16D show the total saccade variability to be expected if the deterministic and stochastic sources combined as independent sources of variability (addition of variances), or if they were perfectly correlated (addition of standard deviations).

In the absence of stochastic noise, the saccade distribution for the normal retina has a mean of 5.97 and a standard deviation of 1.36. For the Alcatraz retina the corresponding values are 5.75 and 2.70. The high variability is related to the long-short strategy discussed earlier.

Clearly, the deterministic and stochastic sources of variability

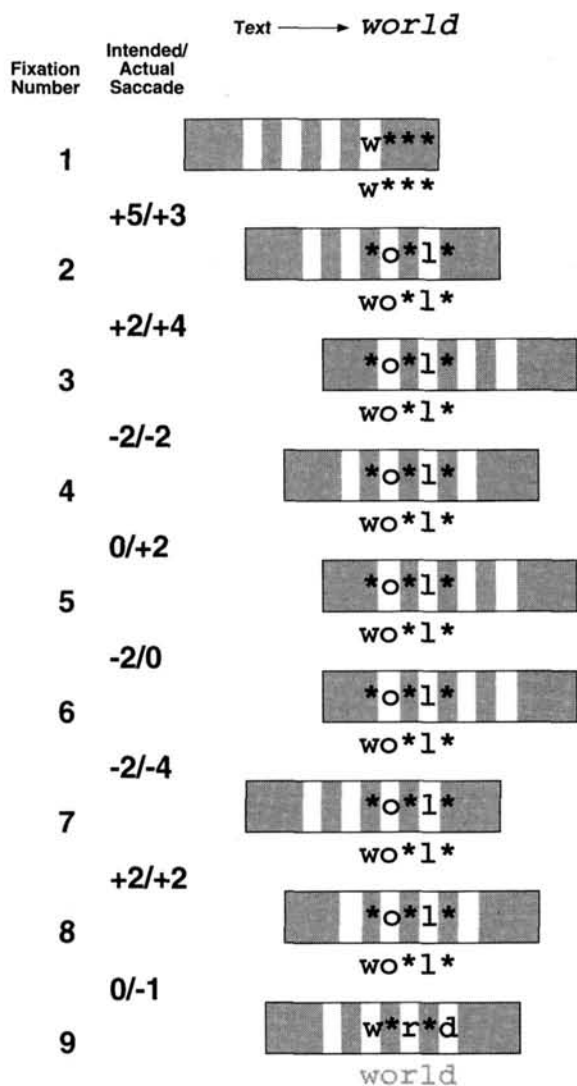


Figure 15. Series of fixations in the presence of saccade noise when Mr. Chips has an Alcatraz retina. The noise had a standard deviation of 1.6, independent of saccade length. The text describes the peculiar behavior of Mr. Chips, including his intention to make zero-length saccades. A sample animation for Figure 15, demonstrating a sequence of Mr. Chips's saccades, is available at the following WorldWideWeb address: <http://vision.psych.umn.edu/chips.html>. * = letters have been detected but not recognized; _ = a space has been recognized.

do not add as do independent sources of noise. Coupling is not surprising because Mr. Chips's algorithm compensates for the effects of noise. For example, a regression that is excessively long because of noise might be followed by an unusually long rightward saccade to make up the lost ground.

There is an important difference between the two retinas for small amounts of saccade noise ($SD < 1.0$). For the unimpaired retina, there is a direct trade-off between total saccade variability and the mean saccade length; their sum has a roughly constant value of 7. The effect of a small amount of noise on mean

saccade length is much more dramatic for the Alcatraz retina; as the saccade noise increases from 0 to 1.0, the mean saccade length drops from 5.75 to 2.95. The reason is that relatively small amounts of noise disrupt the short saccades that are critical to success in the short-long strategy used with the Alcatraz retina. A retinal placement that is offset by one position is much more devastating to the Alcatraz retina than to a normal retina. The peak in saccade variability for the Alcatraz retina (noise SD in the range 0.6 to 0.8 in Figure 16D) occurs when a growing number of regressions intrude in the long-short strategy to combat saccade noise. At even higher noise levels, the long-short strategy is entirely washed out.

Discussion

Is Mr. Chips a Model of Human Reading?

Ideal-observer models are not necessarily intended to model human performance. An ideal-observer model pinpoints performance that is limited by the explicit informational constraints inherent in the task. When human performance matches ideal performance, it is likely that human behavior is also limited by (i.e., explained by) the same constraints. Visual intensity discrimination is a classic example. An ideal observer, whose intensity-discrimination performance is limited by photon noise in the stimulus, obeys the Rose-DeVries law, $\Delta I \propto \sqrt{I}$, where ΔI is the increment threshold and I is the target intensity. For some stimulus conditions at low light levels, human intensity discrimination obeys the same square-root law (cf. Shapley & Enroth-Cugell, 1984). At higher light levels, human performance obeys Weber's law, $\Delta I \propto I$. The prevailing interpretation is that human intensity discrimination is quantum limited in the square-root region at low light levels, but is limited by physiological processes at higher light levels. In this example, comparison with the ideal observer (a) explains human discrimination performance at low light levels and (b) pinpoints stimulus conditions in which performance is limited by perceptual mechanisms.

Mr. Chips shows how humans would behave if their performance were limited strictly by the explicit informational constraints present in the simple reading task. Some aspects of Mr. Chips's behavior are qualitatively similar to human behavior. These similarities suggest that the corresponding human behaviors are limited by the same informational constraints.

We are not proposing that humans do a complex calculation such as entropy minimization in reading. Mr. Chips is obligated to perform such complex calculations to optimize performance. One of our striking findings is that nearly optimal performance can sometimes be achieved with very little computation (e.g., behavior of Chips Jr. and the purely visual observer in Figure 12 and the use of heuristic strategies in Table 4). These examples make clear that nearly ideal human performance does not necessarily imply that humans try to do the same calculations as the ideal observer. One characteristic of human adaptability may be skill in deriving simple heuristics to approximate the computationally intensive performance of ideal observers.

In short, the intended contribution of Mr. Chips is in making

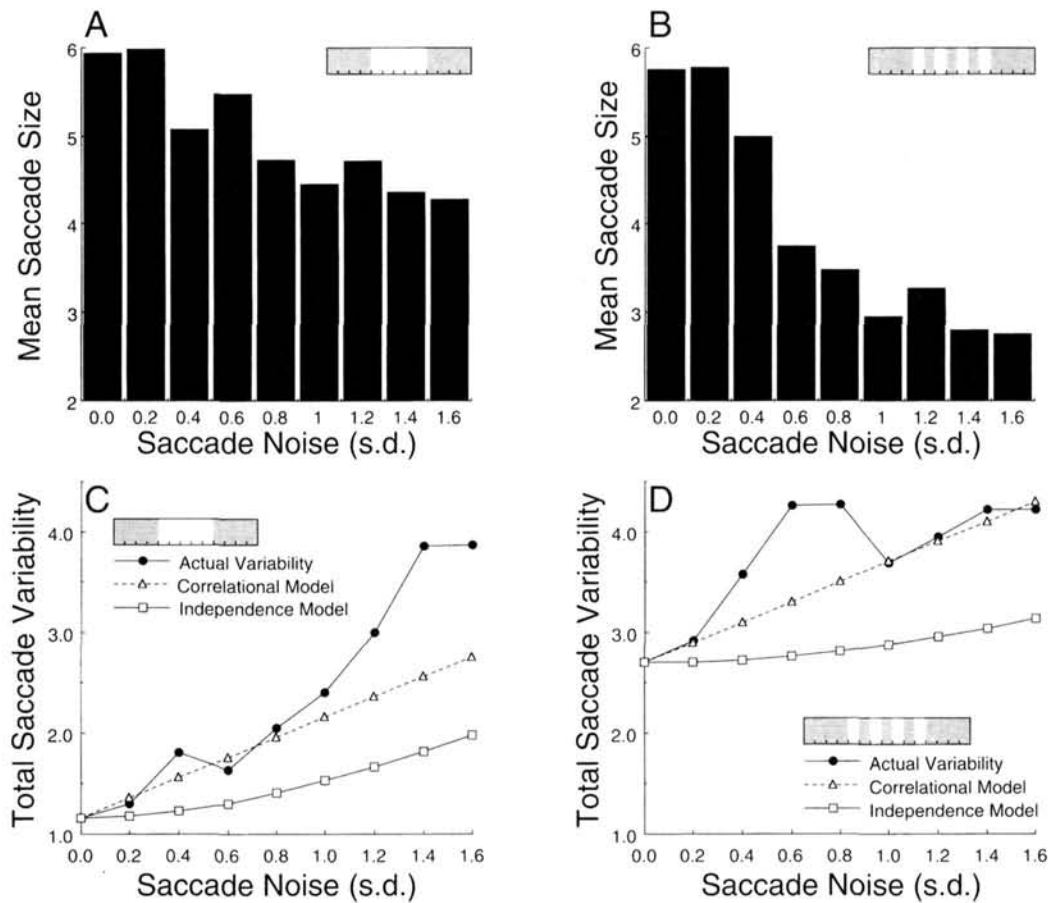


Figure 16. Effect of different levels of saccade noise on mean saccade length and the total saccade variability (i.e., standard deviation of the saccade distribution). The horizontal axis is the standard deviation of the additive Gaussian saccade noise. Figures 16A and 16C are for a normal retina, and Figures 16B and 16D are for an Alcatraz retina. Curves in Figures 16C and 16D show the total saccade variability to be expected if the deterministic and stochastic sources combined as independent sources of variability (addition of variances) or if they were perfectly correlated (addition of standard deviations).

explicit the informational constraints inherent in reading and the behavior that can be accounted for by those constraints.

Mr. Chips exists in a narrowly constrained but well-defined framework. We conclude this article by discussing three theoretical generalizations: fixation linking, elaboration of the retinal model, and use of a more global optimizing principle.

The Fixation-Linking Problem

When there is saccade noise, Mr. Chips needs some way to figure out where his gaze ends up after a saccade. Fixation linking is not a part of the current model. Evidence cited earlier for saccade noise in reading suggests that humans too face a fixation-linking problem.

We refer to a pair of successive fixations as F1 and F2. During F1, Mr. Chips uses knowledge of saccade-noise parameters to work out the landing-site probability distributions. When he arrives at F2, he retains knowledge of these probabilities.

Purely visual pattern matching. It is possible to use visual matching, with no reference to letters or words. An icon is encoded from each character slot in F1. (For low-resolution slots, the icon is binary: something or nothing.) For any intended saccade length, say 7, the landing-site probability table gives the range of possible saccade sizes J , say 5–9. For a retina containing N slots (e.g., 10), there is an overlap of $N-J$ elements from F1 to F2. In our example, for a saccade of 7, the right 3 elements viewed by the retina in F1 should match the left 3 elements in F2. For saccades of length N or greater, there is no overlap between F1 and F2, so there is no visual information available for matching. For each possible length J with nonzero overlap, we check for consistency between what we find in the overlapping slots of F2 and the corresponding slots of F1. This is a form of cross correlation after shifting F1 by J .

This process is guaranteed to find one or more saccade sizes in which there is a match between F1 and F2. If there is just one saccade size yielding a match, the fixation-linking problem

is solved unambiguously. If there are two or more values of J yielding successful matches, the landing-site probability distribution gives the prior probabilities of the alternatives.

How effective is this purely visual strategy? Ambiguity results when there is an accidental match in all overlapping slots for an incorrect saccade size. The probability of an accidental match depends on the amount of overlap, the spatial resolution (high or low) of the retinal slots participating in the overlap, and the redundancy of the text symbols. If the overlap involves only low-resolution slots (such as the right periphery on F1 and the left periphery on F2), the probability of accidental matches is much higher than if high-resolution slots are involved.

There is evidence against purely visual matching as a mechanism for fixation linking in human reading (see Rayner & Pollatsek, 1989, chap. 4). For instance, facilitation due to parafoveal preview survives a change in case (upper to lower) across fixations (Rayner, McConkie, & Zola, 1980).

Visual-lexical matching. In the spirit of the ideal reader, all available information could be used in fixation linking. The purely visual strategy needs to be amended in two ways. First, where possible, letters must be identified; iconic representation without labels will not do for lexical analysis. Second, lexical constraints must be used to reduce the probability of accidental matches and to compute the odds of ambiguous alternatives.

The purely visual and visual-lexical matching strategies both rely on overlap between F1 and F2. For long saccades in which there is no overlap, only the landing-site probabilities are available for linking. There is no guarantee of a correct solution. One way of dealing with this problem would be to regress to a location for which the overlap was sufficient to eliminate ambiguity. Another approach would be to carry ambiguous alternatives along in a set of branching analyses. Perhaps "wrong" branches would reach an inconsistency and could be pruned. An ideal reader might end up with two or more separate interpretations of the text string, only one of which is guaranteed to be correct. Finally, an ideal reader might live with some error in linking. An upper bound on the length of an intended saccade could be adopted to keep the probability of an error below a criterion level.

Fixation linking gives a prominent role to the left side of the visual field. As described earlier, Guez et al. (1993) found that many of their patients selected PRLs to the left of their central scotomas. The authors speculated that this choice of PRL was motivated by fixation linking.

Generalizing the Retinal Model

One way of generalizing Mr. Chips's retinal model is to allow finer gradations of partial information through the retinal slots. Suppose that low-level sensory mechanisms compute a vector of letter probabilities—here termed a *P vector*—for each of Mr. Chips's retinal slots. Each *P* vector contains the likelihoods of the observed image data, given the 27 possible symbols (26 lowercase letters and a space). For example, the vector \mathbf{P}^j for the j th slot would be $[p(a), p(b), \dots, p(z), p(_)]$ where $p(a)$ is the probability of the image data in the j th slot given that the stimulus letter is a , etc., and $p(_)$ is the probability of image data given a space.

Mr. Chips could use the *P* vectors along with his lexical knowledge to compute the probabilities of all possible candidate words. In a straightforward generalization of the algorithm described earlier, these probabilities could be used to narrow the set of candidate words and to compute entropy.

The notion of *P* vectors provides a way for framing the debate between letter- and word-based models of word recognition. *P* vectors constitute a representation of least commitment for the sensory front end. Assuming the *P* vectors are computed ideally, they retain all the available image information pertinent to word recognition. If subsequent processing makes ideal use of lexical constraints, then the system as a whole is an ideal word recognizer. Such a system is equivalent to an ideal observer that does recognition on the basis of a set of neural templates for words rather than letters.¹³

Although Mr. Chips may be able to use *P* vectors, there is evidence that humans do not. Humans appear to commit to letter identity prior to lexical analysis. Pelli et al., (1995) showed that sets of 26 common 3- and 5-letter words have approximately the same contrast thresholds for recognition as the 26 letters of the alphabet. They reasoned that if probabilistic information was available about letters (i.e., information akin to *P* vectors), the contrast thresholds for recognizing words should have been lower than the thresholds for letters. Pelli et al. (1995) interpreted their data to mean that humans recognize letters first and then base word recognition on identified letters.

A system committed to letter identification before word recognition can be clever and learn from mistakes. Over a long series of trials, the system can build a confusion matrix. The cells of the confusion matrix specify the average probabilities of stimulus letters L_i when the response is L_j .¹⁴ Pelli et al. (1995) suggest that the word-superiority effect (Reicher, 1969) can be explained by supposing that participants can make use of these confusion matrices to identify letters in the context of words, but not for isolated letters.

Neural-network models of word recognition, such as the influential *interactive-activation model (IAM)* of McClelland and Rumelhart (1981), blur the distinction between letter and word

¹³ Here, we prove the equivalence of a word-based ideal word recognizer and one that uses *P* vectors containing letter probabilities. An ideal word recognizer chooses the word $_word_$ from its lexicon so that $p(_image_ _word_)p(_word_)$ is maximized. Here, $p(_word_)$ represents the probability of the word in the lexicon. For simplicity, consider the case of a 3-letter word. The image consists of three slots (S_1, S_2, S_3), and any particular word has three letters (L_1, L_2, L_3). So, $p(_image_ _word_)$ can be represented as $p(S_1, S_2, S_3 | L_1, L_2, L_3)$. Assuming the noise process is independent in each slot, $p(S_1, S_2, S_3 | L_1, L_2, L_3) = p(S_1 | L_1, L_2, L_3)p(S_2 | L_1, L_2, L_3)p(S_3 | L_1, L_2, L_3)$. Moreover, assuming no interaction between the slots in image formation, the assumed letter identity in one slot does not affect image formation in another slot, $p(S_1 | L_1, L_2, L_3)$ is the same as $p(S_1 | L_1)$, and likewise for the other slots. Putting everything together, we have $p(_image_ _word_)p(_word_) = p(S_1 | L_1)p(S_2 | L_2)p(S_3 | L_3)p(_word_)$. For any slot, say Slot 1, the probabilities $p(S_1 | L_1)$ for all 27 possible L1s are the elements of a *P* vector.

¹⁴ The rows of a confusion matrix for letters (stimuli on the rows and responses on the columns) represent average *P* vectors associated with the stimulus letters.

recognition. An appropriate selection of connection weights in the general IAM architecture could probably yield ideal word-recognition performance. The specific architecture of the Rumelhart and McClelland model departs from the ideal. This is not because it commits to letter identification prior to word recognition; before the word-to-letter feedback sets in, the activity at the letter level resembles the information in **P** vectors. Subideal performance results from their use of uniform connection weights that would not encode differential probabilities of letters or words. We also note that IAM differs from Mr. Chips in two fundamental ways: (a) It contains a representation of letter features, and (b) it does not contain an eye-movement process for moving from word to word.

The model embodied in Mr. Chips, and the discussion of **P** vectors in this section, both assume that early sensory processing of letters is independent from slot to slot. Even if we accept the notion of discrete slots, there may be interactions between the slots. Evidence for lateral masking indicates that this is likely, especially in peripheral vision.

Incorporation of lateral masking into Mr. Chips would require a detailed model of the interaction. Arditi (1994) has suggested that spatial-frequency filtering could produce a spread of neural excitation from one letter impinging on adjacent letters. In principle, a linear interaction of this sort could be taken into account in the ideal computation of **P** vectors. Instead of keeping a set of 27 neural templates for each retinal slot, Mr. Chips would keep $27^3 = 19,683$ neural templates, representing all possible trigrams centered at that slot. (Mr. Chips could achieve some economy by deleting all of the trigrams that could never occur, given his lexicon.) Using this approach, an ideal observer could nearly eliminate all effects on performance of linear, lateral interactions (Tjan, Legge, & Braje, 1995).

Global Relatives of Mr. Chips

As formulated in this paper, Mr. Chips's saccade planning is local in space because saccade planning focuses on identification of a single word. It is local in time because only a single saccade is planned at a time. We can generalize the model to a family of models, varying in how global they are in space and time, but all with the same computational structure.

Hypersaccades. First, we consider generalizing the model to multiple saccades. If Mr. Chips is working on a given word, there is a finite number of possible placements of his retina, N , that leave his retina in contact with the word. There are many possible sequences of saccades to visit all these retinal placements in exploring the word. We refer to each of these sequences as a *hypersaccade*.

Mr. Chips's ability to plan hypersaccades may be constrained in two ways. First, he may be limited to planning a maximum of n saccades at a time. For the version of Mr. Chips described in this article, $n = 1$. If n is small, there may be no hypersaccade that guarantees word identification. If so, the generalized Mr. Chips chooses the hypersaccade that minimizes expected uncertainty about the word. This is a direct generalization of the algorithm described in this article.

If $n > 1$, a new question arises. Is the entire hypersaccade executed automatically before more saccade planning, or is the

hypersaccade updated after one or more steps of the sequence? (Presumably, the sequence terminates once the word is identified.) Let n' be the number of saccades executed before a new optimal hypersaccade is planned. Best performance should occur if $n' = 1$, that is, a new optimal hypersaccade is computed after each individual saccade, taking into account whatever new visual information has become available.

There is evidence that people preprogram sequences of saccades for scanning geometrical patterns (Zingale & Kowler, 1987). Morrison (1984) has proposed that more than one reading saccade can be programmed during a single fixation. But in his model, multiple-saccade planning results from a mismatch between the times required for word recognition and saccade programming and is not part of a strategic process.

Hyperwords. Similarly, we can generalize Mr. Chips to perform a more global analysis in space. He can analyze more than one word at a time. Assuming the text is of finite length and is composed of words from a finite lexicon, there is a maximum number M of possible word sequences. We refer to each of these M possible sequences of words as an *hyperword*. Of course, in the most general case, the number of hyperwords in Mr. Chips's "hyperlexicon" would be prohibitively large. For a 20-word text and a 50-word lexicon, the number of possible hyperwords would be 50^{20} .

By analogy to the constraints on hypersaccades, we may limit the number of words m that Mr. Chips can analyze as a chunk. The simulation results of this article pertain to $m = 1$. If $m = 2$, Mr. Chips computes hypersaccades for two words at a time. His lexicon consists of hyperwords and their probabilities, composed of all possible pairs of words. To expand the 542-word lexicon, we would need to compile probabilities for the $542^2 = 293,763$ biwords. If these probabilities were compiled from written English text, many would be zero, reflecting syntactic and semantic constraints. We can imagine generalizing Mr. Chips to handle hyperwords of $m = 2$ or $m = 3$, but not much more.

By analogy to saccade planning, Mr. Chips could update the hyperword under study when one or more of the constituent words is identified. Call this parameter m' . For example, if $m = 3$ and $m' = 2$, Mr. Chips would analyze texts in chunks of three words. When at least two of the three words are identified, he would set his sites on a new hyperword (which would contain the unidentified word from the previous hyperword and two additional words).

It is possible that humans plan reading saccades on the basis of chunks of text larger than a single word, perhaps phrases. The number of words in phrases is not constant, so some flexibility in the hyperword concept would be necessary to accommodate such behavior.

The extended chips family. The members of the extended Chips family of models can be characterized by four parameters: n is the number of saccades planned at one time (length of hypersaccades); n' is the number of saccades executed in a sequence before a new hypersaccade is planned; m is the number of words analyzed in parallel (size of hyperwords); and m' is the number of words recognized before a new hyperword is selected for study.

For all members of the extended Chips family, the same generalized optimization principle applies:

Make the hypersaccade that minimizes expected entropy about the hyperword.

References

- Akutsu, H., Legge, G. E., Ross, J. A., & Schuebel, K. J. (1991). Psychophysics of reading: X. Effects of age-related changes in vision. *Journal of Gerontology: Psychological Sciences*, *46*, 325–331.
- Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471–517.
- Arditi, A. (1994). On the relationship of letter acuity to reading acuity. In A. C. Kooijman, P. L. Looijestijn, J. A. Welling, & G. J. van der Wildt (Eds.), *Low vision: Research and new developments in rehabilitation* (pp. 38–45). Amsterdam: IOS Press.
- Arditi, A., Knoblauch, K., & Grunwald, I. (1990). Reading with fixed and variable character pitch. *Journal of the Optical Society of America A*, *7*, 2011–2015.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.
- Bailey, I., & Lovie, J. (1980). The design and use of a new near vision chart. *American Journal of Optometry and Physiological Optics*, *57*, 378–387.
- Baldasare, J., Watson, G. R., Whittaker, S. G., & Miller-Shaffer, H. (1986). The development and evaluation of a reading test for low vision individuals with macular loss. *Journal of Visual Impairment and Blindness*, *80*, 785–789.
- Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Research*, *18*, 637–650.
- Barlow, H. B., & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, *19*, 783–793.
- Beckmann, P. J., & Legge, G. E. (1996). Psychophysics of reading: XIV. The page-navigation problem in using magnifiers. *Vision Research*, *36*, 3723–3733.
- Bennett, P. J., & Banks, M. S. (1991). The effects of contrast, spatial scale, and orientation on foveal and peripheral phase discrimination. *Vision Research*, *31*, 1759–1786.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178.
- Braje, W. L., Tjan, B. S., & Legge, G. E. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research*, *35*, 2955–2966.
- Bullimore, M. A., & Bailey, I. (1995). Reading and eye movements in age-related maculopathy. *Optometry and Vision Science*, *72*, 125–138.
- Burgess, A., & Barlow, H. B. (1983). The precision of numerosity discrimination in arrays of random dots. *Vision Research*, *23*, 811–820.
- Caramazza, A., & Hillis, A. E. (1990). Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature*, *346*, 267–269.
- Carpenter, P. A., & Just, M. A. (1977). Reading comprehension as eyes see it. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 109–139). Hillsdale, NJ: Erlbaum.
- Carroll, J. B., Davies, P., & Richman, B. (1971). *The American Heritage word frequency book*. Boston: Houghton Mifflin.
- Carver, R. P. (1990). *Reading rate: A review of research and theory*. San Diego: Academic Press.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Coeffe, C., & O'Regan, J. K. (1987). Reducing the influence of non-target stimuli on saccade accuracy: Predictability and latency effects. *Vision Research*, *27*, 227–240.
- Cummings, R. W., Whittaker, S. G., Watson, G. R., & Budd, J. M. (1985). Scanning characters and reading with a central scotoma. *American Journal of Optometry and Physiological Optics*, *62*, 833–843.
- Dickinson, C., & Rabbitt, P. M. A. (1991). Simulated visual impairment: Effects on text comprehension and reading speed. *Clinical Vision Sciences*, *6*, 301–308.
- Epelboim, J., Booth, J. R., & Steinman, R. M. (1994). Reading unspaced text: Implications for theories of reading eye movements. *Vision Research*, *34*, 1735–1766.
- Faye, E. E. (Ed.). (1984). *Clinical low vision*. Boston: Little, Brown.
- Ferris, F. L. (1983). Senile macular degeneration: Review of epidemiologic features. *American Journal of Epidemiology*, *118*, 132–149.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, 381–391.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*, 267–314.
- Goodrich, G. L., & Quillman, R. D. (1977). Training eccentric viewing. *Journal of Visual Impairment and Blindness*, *71*, 377–381.
- Guez, J., le Gargasson, J., Rigaudiere, F., & O'Regan, J. K. (1993). Is there a systematic location for the pseudo-fovea in patients with central scotoma? *Vision Research*, *33*, 1271–1279.
- He, P., & Kowler, E. (1991). Saccadic localization of eccentric forms. *Journal of the Optical Society of America A*, *8*, 440–449.
- Hecht, S., Shlaer, S., & Pirenne, M. H. (1942). Energy, quanta, and vision. *Journal of General Physiology*, *25*, 819–840.
- Heinen, S. J., & Skavenski, A. A. (1992). Adaptation of saccades and fixation to bilateral foveal lesions in adult monkey. *Vision Research*, *32*, 365–373.
- Hess, R. F., & Hayes, A. (1994). The coding of spatial position by the human visual system: Effects of spatial scale and retinal eccentricity. *Vision Research*, *34*, 625–643.
- Hilton, J. (1934). *Goodbye, Mr. Chips*. Boston: Little, Brown.
- Hyona, J. (1995). Do irregular letter combinations attract readers' attention? Evidence from fixation locations in words. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 68–81.
- Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General*, *108*, 151–181.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.
- Kemper, S. (1992). Language and aging. In F. I. M. Craik & T. A. Salt-house (Eds.), *The handbook of aging and cognition* (pp. 213–270). Hillsdale, NJ: Erlbaum.
- Kersten, D. (1990). Statistical limits to image understanding. In C. Blakemore (Ed.), *Vision: Coding and efficiency*. Cambridge, England: Cambridge University Press.
- Klein, S. A., & Levi, D. M. (1987). The position sense of the peripheral retina. *Journal of the Optical Society of America A*, *4*, 1544–1553.
- Klitz, T. S., & Legge, G. E. (1994). Modeling the visual span in reading [Abstract]. *Investigative Ophthalmology and Visual Science*, *35*(Suppl. 4), 1949.
- Klitz, T. S., Mansfield, J. S., & Legge, G. E. (1995). Reading speed is affected by font transitions [Abstract]. *Investigative Ophthalmology and Visual Science*, *36*(Suppl. 4), S670.
- Kowler, E., & Blaser, E. (1995). The accuracy and precision of saccades to small and large targets. *Vision Research*, *35*, 1741–1754.

- Legge, G. E. (1992). An ideal observer model of reading [Abstract]. *Investigative Ophthalmology and Visual Science*, 33(Suppl. 4), 1414.
- Legge, G. E. (1995). Mr. Chips (Version 4.3) [Computer software]. Minneapolis, MN: Author.
- Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1995). *Psychophysics of reading: XVI. The visual span in normal and low vision*. Manuscript submitted for publication.
- Legge, G. E., Gu, Y., & Luebker, A. (1989). Efficiency of graphical perception. *Perception & Psychophysics*, 46, 365–374.
- Legge, G. E., Parish, D. H., Luebker, A., & Wurm, L. H. (1990). Psychophysics of reading: XI. Comparing color contrast and luminance contrast. *Journal of the Optical Society of America A*, 7, 2002–2010.
- Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading: I. Normal vision. *Vision Research*, 25, 239–252.
- Legge, G. E., Ross, J. A., Isenberg, L. M., & LaMay, J. M. (1992). Psychophysics of reading: XII. Clinical predictors of low-vision reading speed. *Investigative Ophthalmology and Visual Science*, 33, 677–687.
- Legge, G. E., Ross, J. A., Luebker, A., & LaMay, J. M. (1989). Psychophysics of reading: VIII. The Minnesota low-vision reading test. *Optometry and Vision Science*, 66, 843–853.
- Legge, G. E., & Rubin, G. S. (1986). Psychophysics of reading: IV. Wavelength effects in normal and low vision. *Journal of the Optical Society of America A*, 3, 40–51.
- Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading: V. The role of contrast in normal vision. *Vision Research*, 27, 1165–1171.
- Legge, G. E., Rubin, G. S., Pelli, D. G., & Schleske, M. M. (1985). Psychophysics of reading: II. Low vision. *Vision Research*, 25, 253–266.
- Lima, S. D. (1993). Word-initial letter sequences and reading. *Current Directions in Psychological Science*, 2, 139–142.
- Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, 35, 549–568.
- Mansfield, J. S., Ahn, S. J., Legge, G. E., & Luebker, A. (1993). A new reading-acuity chart for normal and low vision: Ophthalmic and visual optics/noninvasive assessment of the visual system. *Optical Society of America Technical Digest*, 3, 232–235.
- Mansfield, J. S., Legge, G. E., & Bane, M. (1996). Psychophysics of reading: XV. Font effects in normal and low vision. *Investigative Ophthalmology and Visual Science*, 37, 1492–1501.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375–407.
- McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye-movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, 28, 1107–1118.
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 16, 578–586.
- McMahon, T. T., Hansen, M., Stelmack, J., Oliver, P., & Viana, M. A. (1993). Saccade eye movements as a measure of the effect of low vision rehabilitation on reading rate. *Optometry and Vision Science*, 70, 506–510.
- Miller, G. A., & Selfridge, J. A. (1950). Verbal context and recall of meaningful material. *American Journal of Psychology*, 63, 176–185.
- Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 667–682.
- Morrison, R. E., & Rayner, K. (1981). Saccade size in reading depends on character spaces and not visual angle. *Perception & Psychophysics*, 30, 395–396.
- Morton, J. (1964). The effects of context upon speed of reading, eye movements and eye-voice span. *Quarterly Journal of Experimental Psychology*, 16, 340–354.
- Nilsson, U. L. (1990). Visual rehabilitation with and without educational training in the use of optical aids and residual vision: A prospective study of patients with advanced age-related macular degeneration. *Clinical Vision Sciences*, 6, 3–10.
- O'Regan, J. K. (1990). Eye movements and reading. In E. Kowler (Ed.), *Eye Movements and their role in visual and cognitive processes* (pp. 395–453). New York: Elsevier Science.
- O'Regan, J. K. (1991). Understanding visual search and reading using the concept of stimulus "grain." *IPO Annual Progress Report*, 26, 96–108.
- O'Regan, J. K., Levy-Schoen, A., & Jacobs, A. M. (1983). The effect of visibility on eye-movement parameters in reading. *Perception & Psychophysics*, 34, 457–464.
- Parish, D. H., & Legge, G. E. (1988). The effect of jitter on reading rate and visual acuity [Abstract]. *Investigative Ophthalmology and Visual Science*, 29(Suppl. 3), 141.
- Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, 31, 1399–1415.
- Peli, E. (1986). Control of eye movements with peripheral vision: Implications for training of eccentric viewing. *American Journal of Optometry and Physiological Optics*, 63, 113–118.
- Pelli, D. G., Burns, C. W., Farrell, B., & Moore, D. C. (in press). Identifying letters. *Vision Research*.
- Pelli, D. G., Farrell, B., & Moore, D. C. (1995). *The role of letters in recognizing words*. Manuscript submitted for publication.
- Ramachandran, V. S. (1992). Filling in gaps in perception: I. *Current Directions in Psychological Science*, 1, 199–205.
- Ramachandran, V. S. (1993). Filling in gaps in perception: II. Scotomas and phantom limbs. *Current Directions in Psychological Science*, 2, 56–65.
- Rayner, K. (1978). Foveal and parafoveal cues in reading. In J. Requin (Ed.), *Attention and performance: VII* (pp. 149–161). Hillsdale, NJ: Erlbaum.
- Rayner, K. (1993). Eye movements in reading: Recent developments. *Current Directions in Psychological Science*, 2, 81–85.
- Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science*, 206, 468–469.
- Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, 16, 829–837.
- Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology*, 12, 206–226.
- Rayner, K., & Morris, R. K. (1992). Eye movement control in reading: Evidence against semantic preprocessing. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 163–172.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Perception & Psychophysics*, 31, 337–350.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275–280.
- Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, 37, 495–510.
- Rubin, G. S., & Legge, G. E. (1989). Psychophysics of reading: VI. The role of contrast in low-vision. *Vision Research*, 29, 79–91.

- Rubin, G. S., & Turano, K. (1994). Low vision reading with sequential word presentation. *Vision Research*, *34*, 1723–1733.
- Schlingensiepen, K.-H., Campbell, F. W., Legge, G. E., & Walker, T. D. (1986). The importance of eye movements in the analysis of simple patterns. *Vision Research*, *26*, 1111–1117.
- Schuchard, R. A., & Fletcher, D. (1994). Preferred retinal locus: A review with applications in low vision rehabilitation. *Ophthalmology Clinics of North America*, *7*, 243–255.
- Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, *30*, 50–64.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. In N. Osborne & G. Chader (Eds.), *Progress in retinal research* (Vol. 3, pp. 263–346). Oxford, England: Pergamon Press.
- Sloan, L. L. (1977). *Reading aids for the partially sighted: A systematic procedure for prescribing reading aids*. Baltimore: Williams & Wilkins.
- Snellen (1864). *Test-types for the determination of the acuteness of vision* (2nd ed.). Utrecht, The Netherlands: P. W. van de Weijer.
- Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, *369*, 395–397.
- Spragins, A. B., Lefton, L. A., & Fisher, D. F. (1976). Eye movements while reading and searching spatially transformed text: A developmental examination. *Memory & Cognition*, *4*, 36–42.
- Timberlake, G. T., Mainster, M. A., Peli, E., Augliere, R. A., Essock, E. A., & Arend, L. E. (1986). Reading with a macular scotoma: I. Retinal location of scotoma and fixation area. *Investigative Ophthalmology and Visual Science*, *27*, 1137–1147.
- Timberlake, G. T., Peli, E., Essock, E. A., & Augliere, R. A. (1987). Reading with a macular scotoma: II. Retinal locus for scanning text. *Investigative Ophthalmology and Visual Science*, *28*, 1268–1274.
- Tinker, M. A. (1963). *Legibility of print*. Ames: Iowa State University Press.
- Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, *35*, 3053–3069.
- Tjan, B. S., Legge, G. E., & Braje, W. L. (1995). Quantifying stimulus information for recognition [Abstract]. *Investigative Ophthalmology and Visual Science*, *36*(Suppl. 4), S670.
- Trauzettel-Klosinski, S., Teschner, C., Tornow, R.-P., & Zrenner, E. (1994). Reading strategies in normal subjects and in patients with macular scotoma assessed by two new methods of registration. *Neuro-ophthalmology*, *14*, 15–30.
- Turbo-C (Version 1.5) [Computer programming language]. (1987). Scotts Valley, CA: Borland International.
- Underwood, G., Clews, S., & Everatt, J. (1990). How do readers know where to look next? Local information distributions influence eye fixations. *Quarterly Journal of Experimental Psychology*, *42A*, 39–65.
- van Opstal, A. J., & van Gisbergen, J. A. M. (1989). Scatter in the metrics of saccades and properties of the collicular motor map. *Vision Research*, *29*, 1183–1196.
- Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification function. *Experimental Brain Research*, *37*, 475–494.
- White, J., & Bedell, H. (1990). The oculomotor reference in humans with bilateral macular disease. *Investigative Ophthalmology and Visual Science*, *31*, 1149–1161.
- Whittaker, S. G., Cummings, R., & Swieson, L. (1991). Saccade control without a fovea. *Vision Research*, *31*, 2209–2218.
- Whittaker, S. G., & Lovie-Kitchin, J. (1993). Visual requirements for reading. *Optometry and Vision Science*, *70*, 54–65.
- World Health Organization. (1966). Blindness information collected from various sources. *Epidemiology and Vital Statistics Report*, *19*, 427–511.
- Zingale, C. M., & Kowler, E. (1987). Planning sequences of saccades. *Vision Research*, *27*, 1337–1341.

Received October 12, 1995

Revision received August 2, 1996

Accepted August 12, 1996 ■