PERGAMON

Vision Research 42 (2002) 2219–2234

Vision Research

www.elsevier.com/locate/visres

Mr. Chips 2002: new insights from an ideal-observer model of reading

Gordon E. Legge [a,*], Thomas A. Hooven [a], Timothy S. Klitz [b], J. Stephen Mansfield [c], Bosco S. Tjan [d]

[a] Minnesota Laboratory for Low-Vision Research, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455, USA

[b] Department of Psychology, Washington and Jefferson College, USA

[c] Department of Psychology, Plattsburgh State University, USA

[d] Department of Psychology, University of Southern California, USA

Received 9 February 2001; received in revised form 25 February 2002

Abstract

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. This model provides a computational framework for interpreting human reading saccades in both normal and low vision. The purpose of this paper is to report performance of the model for conditions emulating reading with normal vision—a visual span of nine characters, multiplicative saccade noise with a standard deviation of 30%, and texts based on three full-length children's books. Comparison of fixation locations by humans and Mr. Chips revealed: (1) that both exhibit very similar word-skipping behavior; (2) both show initial fixations near the center of words, but with a systematic difference suggestive of an asymmetry in the human visual span; and (3) differences in the pattern of refixations within words that may uncover non-optimal lexical inference by human readers. A human context effect—30% difference in mean saccade size between continuous text and random sequences of words—was very similar to the 25% effect for the model associated with a corresponding difference in the predictability of text words. Overall, our findings show that many of the complicated aspects of human reading saccades can be explained concisely by early information-processing constraints.

Keywords: Vision; Reading; Eye movements; Saccades

1. Introduction

Legge, Klitz, and Tjan (1997) described an ideal-observer model of reading. The computer program implementing this model is called *Mr. Chips*. The model is "ideal" in the sense that it uses an optimizing principle: it reads input text sequentially, without error, in the minimum number of saccades. The model's sources of information include visual data, word knowledge, and statistical information about the accuracy of saccadic eye movements. The input texts are strings of words, drawn randomly from a frequency-weighted lexicon. An overview of the model is given below. For brevity, we will refer to the 1997 paper as Chips97.

The major goals of the model were threefold: (1) to examine the impact of low-level informational constraints on reading behavior; (2) to compare relevant aspects of human reading performance to the model's behavior; and (3) to provide a theoretical foundation for understanding the impact of visual-field loss (an important form of visual impairment) on reading by introducing appropriate constraints into the model's visual front-end. Our emphasis on the visual front-end constraints in reading and our reliance on only the lexical information distinguishes our approach from most previous models.

This paper has two major purposes: to generalize the findings of the model by examining behavior for more "natural" text, and to compare its performance with key human studies dealing with fixation locations within words.

* Corresponding author. Tel.: +1-612-625-0846; fax: +1-612-626-2079.

E-mail address: legge@umn.edu (G.E. Legge).

0042-6989/02/$ - see front matter
PII: S0042-6989(02)00131-1

1.1. Overview of the model

Ideal-observer analysis identifies optimal performance on a task, given a specified set of information sources. An ideal-observer analysis is not a model of how humans perform a task, but rather it establishes the pattern of performance to be expected if the available information is used optimally.

Hecht, Shlaer, and Pirenne (1942) introduced the ideal observer to vision with their work on quantum 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 (Chung, Legge, & Tjan, in press; Parish & Sperling, 1991; Pelli, Burns, Farell, & Moore, in press; Solomon & Pelli, 1994), and object recognition (Liu, Knill, & Kersten, 1995; Tjan, Braje, Legge, & Kersten, 1995). In all of these studies, ideal performance is based on optimal use of the information in the stimulus per se, or on the information available following one or more well-specified stages of visual processing, such as image formation and photoreceptor sampling in Geisler's analysis. When human performance matches ideal performance, it can be assumed that humans are using all available information and that their performance limitations are described by the limitations facing the ideal observer. It does not follow, however, that humans must be using the same algorithm as the ideal observer, since more than one algorithm may achieve the same computational goal (cf. Marr, 1982). When humans under-perform the ideal observer, characteristics of the discrepancy may suggest the nature of the additional processing constraints facing the human. For example, deviation of human intensity discrimination from the ideal square-root law (Rose-deVries law) has been instructive in guiding the development of models of light adaptation. It can even happen that humans outperform a specific "ideal" observer. Optimality of the ideal observer with respect to the information explicitly assumed in its formulation implies that humans must be using information unavailable to the ideal observer. This technique has been used to rule out an entire class of theories that suggested object recognition could be achieved with a pure 2-D strategy (Liu et al., 1995).

Mr. Chips is an extension of ideal-observer analysis to a simple reading task. Like the ideal observers in the cited studies, Mr. Chips is not proposed as a model of human behavior, and is not falsifiable by human reading data. Its value in studying human reading should be judged on its claim to optimality (see Chips97), the reasonableness of its assumed informational constraints, and the insights it generates into human reading. When Mr. Chips' behavior is parametrically similar to human reading behavior, it is reasonable to propose that the information constraints explaining Mr. Chips' behavior also explain corresponding human behavior. Under these conditions, human behavior is optimal, implying the use of an algorithm that is equivalent to or that closely approximates the performance resulting from ideal computation for the conditions in question. When there is a discrepancy between the behavior of Mr. Chips and humans, it is reasonable to ask what process in human reading is not captured in the ideal-observer analysis.

Mr. Chips makes optimal use of three sources of information: visual information, lexical information and information about eye-movement accuracy. Visual information is obtained through a "retina" (Fig. 1, top left). The retina is a linear array of character slots. Each slot can either be high resolution (individual letters can be identified) or low resolution (only a distinction between letters and spaces is possible). In Fig. 1, the "normal retina" has nine high-resolution slots, flanked by peripheral regions of four low-resolution slots. Fig. 1 also illustrates a retina with three low-resolution slots in central vision, simulating a central scotoma. For discussion of the model's performance with central scotomas (see Klitz, Legge, & Tjan, 2000; Legge et al., 1997).

We will refer to the number of adjacent, high-resolution slots as the model's visual span. This term has also been used to refer to the number of adjacent letters recognizable on each fixation in human reading (Legge, Ahn, Klitz, & Luebker, 1997; Legge, Mansfield, & Chung, 2001; O'Regan, 1990, 1991). [1] In Chips97, we studied the dependence of Mr. Chips' performance on the size of the visual span. For the computer simulations in this paper, the visual span was kept constant at nine letters, as illustrated in Fig. 1. This value was chosen to match a consensus estimate of the size of the human visual span (see Legge et al., 1997).

The model's visual span has a rectangular profile within which letters are recognized with perfect accuracy. Human visual-span profiles show a more gradual decline in letter accuracy outward from the middle

[1] The notion of "visual span" differs from the concept of "perceptual span" (McConkie & Rayner, 1975). The size of the "perceptual span" depends on factors in addition to letter recognition. 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 four characters to the left.

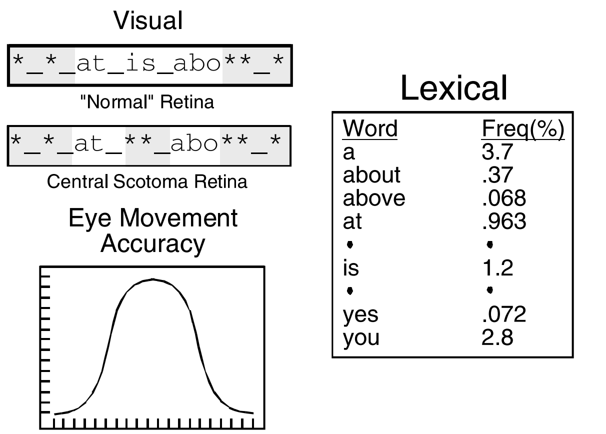

Fig. 1. Three sources of information available to Mr. Chips. See the text for details.

(Legge et al., 2001; Nazir, O'Regan, & Jacobs, 1991; O'Regan, Levy-Schoen, & Jacobs, 1983). It is unknown whether this difference between Mr. Chips and humans accounts for any important discrepancies in their reading behavior.

The second source of information available to Mr. Chips is lexical information (Fig. 1, right). Mr. Chips can use any properly formatted lexicon consisting of words and corresponding frequency information. The simulations reported in Chips97 all used a lexicon composed of the 542 most common words in written English (Carroll, Davies, & Richman, 1971) along with their frequencies of occurrence. One purpose of the present paper is to generalize the model by using larger and more natural lexicons.

The model's third source of information is statistical knowledge of saccade accuracy (Fig. 1, bottom left). Human saccades manifest noise, i.e., random variability in length or direction. Within the model, we have represented this noise as Gaussian variability in saccade length, made discreet so that all saccades are of integer length. The amplitude of the noise is determined by the standard deviation of the Gaussian error. The standard deviation can be kept constant or allowed to scale with saccade size, termed multiplicative noise. The effect of different levels of saccade noise was studied in Chips97. When Mr. Chips plans a saccade, he knows the parameters (mean and standard deviation) of the probability distribution of landing sites. He takes this distribution into account in planning the optimal saccade. For the simulations reported in this paper, the model used multiplicative noise with a standard deviation equal to 30% of saccade length, consistent with empirical estimates for human reading saccades (Legge et al., 1997).

Mr. Chips' text is constructed by drawing words at random from his lexicon according to the 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' task is to read the text in the minimum number of saccades, identifying the words in order without error.

The model operates according to an entropy-minimization principle. It identifies words as it moves from left to right through the text. At any point, it 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 sub-set 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 it is identified unambiguously (i.e., its uncertainty is zero, or less than some specified entropy tolerance).

In Fig. 1, the word 'abo**', visible in the normal retina, is unknown because both "above" and "about" fit the visual information present. Mr. Chips will make a saccade of 10 characters to place the last letter of the word in the first slot in the visual span. From knowledge of the lexicon, Mr. Chips knows that the identity of the fifth letter will resolve the ambiguity. Appendix A provides a simple illustration of the entropy-minimization algorithm. Mr. Chips' use of letter-level information to constrain lexical search is similar to cohort models of human word recognition (cf. Johnson & Pugh, 1994). A related concept in the human literature is *neighborhood*, the number of words that differ from a target word by one letter. Experiments have shown that human word recognition is sensitive to the number and frequency of words in the neighborhood (Andrews, 1989) and that reading is slower for words with large neighborhoods (Pollatsek, Binder, & Perea, 1999).

Mr. Chips' optimization rule results in a sequence of saccades through text that includes occasional regressions (leftward saccades) as well as forward saccades. Regressions can occur when Mr. Chips rightward saccade (computed on the basis of expected entropy reduction) leaves critical information about a word out of view on the left of the visual span. The model's regressions are discussed in detail in Chips97.

It is worth repeating that Mr. Chips is not presented as a model of human performance per se. We do not propose that people do entropy calculations in planning saccades. Instead, the model can be used as a yardstick, indicating the nature of performance if all of the specified information is used optimally. In Chips97, we showed that a much simpler algorithm, based on a small set of heuristic strategies (similar to the "strategy tactics" proposed by O'Regan (1990) and studied by Reilly & O'Regan (1998)), nearly achieve the ideal performance of the model. It is plausible, therefore, that human behavior could nearly match ideal performance while using a computationally simpler algorithm to process the available information.

2. General methods

2.1. Text materials

In addition to the original 542-word lexicon used in Chips97, we used three new text sources—*Grimms' Fairy Tales* (Grimm & Grimm, 1812), *Little Women* (Alcott, 1869), and *Glinda of Oz* (Baum, 1920). These full-length texts were downloaded as text files from Internet sites. Most of the Mr. Chips simulations reported in this paper used the Grimm source, with comparisons made to the other sources as indicated.

Some modifications of the Grimm text were necessary before it could be transformed to a lexicon for Mr. Chips. All detectable typos in the file were corrected. Archaic or British spellings were transformed to their modern American equivalents [2] (e.g., "thou" to "you" and "recognise" to "recognize"). Unnecessary gibberish was deleted (e.g., animal sounds such as "aik aik aik" and song lyrics containing non-words like "dee"). All uppercase letters were transformed to lowercase (Mr. Chips does not observe any distinction). Numbers were converted to their alphabetic equivalent (e.g., 9 to nine). All punctuation was eliminated except for apostrophes. Apostrophes that were used as single quotes were removed, but those used in possessives or in contractions such as can't or won't were preserved. The Mr. Chips algorithm was modified to accept apostrophes as an allowable character in words, along with the 26 letters of the alphabet. Accordingly, the words "its" and "it's" were treated as distinct lexical entries.

Following these transformations, a Perl program was used to create the Grimm lexicon from the Grimm text by identifying all unique words and their relative frequency of occurrence. There were 7504 unique words. Their distribution by frequency ranged from "the" which appeared 19,691 times to 2340 words that appeared only once.

Lexicons were generated in a similar way for *Little Women* and *Glinda of Oz*. Table 1 shows the distribution of words by length in the three books and corresponding values computed for the 542-word lexicon used in Chips97. For comparison, corresponding statistics are shown for the British National Corpus (Kilgarriff, 1997).

Table 1
Word-length statistics from Mr. Chips' four lexical sources and the British National Corpus

Word length (letters)	Percent of words				
	Brothers Grimm	*Glinda of Oz*	*Little Women*	542-word lexicon	British National Corpus
1	3.06	2.77	4.10	3.69	3.15
2	15.79	16.83	16.53	23.49	16.94
3	30.08	25.27	25.48	32.05	21.16
4	22.60	19.97	20.51	22.32	15.90
5	11.12	11.70	11.12	10.77	10.99
6	7.19	8.50	8.39	3.96	8.43
7	4.92	6.51	5.81	2.28	7.95
8	2.88	4.02	3.71	0.64	5.64
9	1.39	2.66	2.09	0.65	4.12
10	0.62	0.99	1.20	0.15	2.71
11	0.23	0.49	0.58	0	1.52
12	0.09	0.19	0.31	0	0.80
13	0.03	0.05	0.13	0	0.44
14	<0.01	0.03	0.03	0	0.17
15	<0.01	0.02	0.01	0	0.06
16	0	0	<0.01	0	0.02
17	0	0	<0.01	0	0.01
18	0	0	<0.01	0	<0.01
19–24	0	0	0	0	<0.01 (each)
Total unique words	7504	3815	11,028	542	74,304
Mean word length	3.93	4.21	4.14	3.42	4.70

2.2. Software implementation and simulation parameters

The original DOS-based Turbo-C version of Mr. Chips was imported to a UNIX SGI platform. The C code was compiled using the SGI C++ compiler (CC). The UNIX version, designated Chips V5.2, could handle much larger lexicons. Simulations were conducted to ensure that the Unix V5.2 version of Mr. Chips yielded the same results for the same input parameters and text as the DOS Turbo-C version. No discrepancies were found.

For all of the simulation results in this paper, Mr. Chips' retina had a visual span of nine high-resolution slots (see Fig. 1) flanked on either side by four low-resolution slots. Except for simulations in which the saccade noise is specified as zero, all the simulations were conducted with multiplicative Gaussian noise. The standard deviation of the noise was 30% of the intended saccade size. For instance, if Mr. Chips planned a sac-

[2] Mr. Chips has no partiality to modern American usage, but potential uses of the same materials in experiments with human subjects in an American laboratory motivated these changes.

cade of 10 letters, a noise sample was added from a Gaussian distribution of 0 mean and standard deviation 3. The actual saccade length was the resulting value, rounded to the nearest integer length. We chose a visual span of nine and multiplicative noise with 30% standard deviation as approximations to corresponding human values (Legge et al., 1997).

For most of the simulations, texts were created by drawing N words at random from the lexicon, with probability proportional to the relative frequency of occurrence in the text source. Mr. Chips then read the text, i.e., processed it with his entropy-minimization algorithm. The sequence of saccades was recorded. Saccade histograms were compiled and the mean and standard deviation were used as summary statistics. Most simulations involved reading 10,000 words. The word-skipping analysis (Fig. 3) used 40,000 words to ensure sufficient sampling of the infrequent longer words. The refixation analysis (Figs. 7 and 8) also used 40,000 because of the low-frequency of refixations.

Although Mr. Chips does not make use of syntactic, semantic, or other inter-word constraints in reading, it is possible that Mr. Chips' saccade behavior could be implicitly responsive to these linguistic constraints. In simulations not reported in detail here we compared Mr. Chips' first-order saccade distributions for "real" and "random" texts. The real texts were continuous passages selected from different stories in *Grimms' Fairy Tales*. There were no systematic differences between the saccade distributions for real and random text. From these results, we conclude that there are no structural characteristics of real text that have a major impact on Mr. Chips' saccade distributions. From the point of view of the model's performance, there is no distinction between reading real text and random text with the same frequency distribution of words.

3. Experiment 1. Lexicon analysis

In Chips97, all of the simulations were conducted with a lexicon consisting of the 542 most common words in written English, based on the compilation of Carroll et al. (1971). This lexicon is small, compared with the number of unique words that might appear in a full-length book, or that is known to a human reader. It contains only high-frequency words and relatively few long words. It is representative of a broad cross-section of texts, but of no text in particular.

The architecture of the model permits use of any arbitrary lexicon. We report here on results for lexicons ranging in size up to 11,028 words from three full-length children's books in addition to the original 542-word Chips lexicon. We focus on two main issues: the effects of lexicon size and context on reading.

3.1. Effect of lexicon size

In Chips97, we showed that mean saccade size depends on lexicon size. We sub-sampled the original 542-word lexicon to show that saccade size grows as lexicon size decreases. In the present paper, we will analyze this relationship in more detail.

The Grimm lexicon of 7504 words was sub-sampled by factors of about 3 to produce six additional sub-lexicons of sizes 10, 31, 91, 278, 834, and 2501 words. The sub-sampling was designed to retain approximately the same frequency distribution and mean word length as the full Grimm lexicon. For each of the sub-lexicons, separate simulations were conducted with and without saccade noise. Each simulation used 2000 words of randomly generated text from the corresponding sub-lexicon.

A similar procedure was used to derive sub-sampled lexicons from *Little Women* and *Glinda of Oz*.

Fig. 2 plots mean saccade size as a function of lexicon size on a log scale. In the upper part of the graph, the solid circles and line are based on *Grimms' Fairy Tales*. The straight line fits most of this range and shows that mean saccade length increases with decreasing lexicon size. The squares and triangles from *Glinda of Oz* and *Little Women* fall close to the same line; there is not much variation across literary sources.

The lower part of the figure shows simulation results when Mr. chips had to contend with multiplicative saccade noise (s.d. = 30% of saccade length). This curve is displaced downward from the no noise case so that

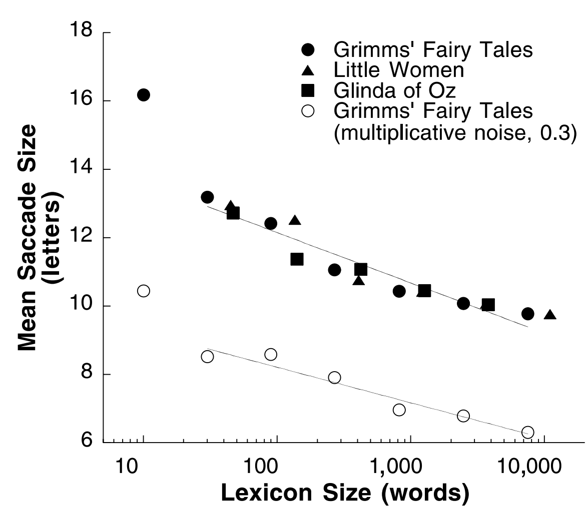

Fig. 2. Mean saccade size as a function of lexicon size for a set of 18 lexicons derived from the *Grimm*, *Glinda*, and *Little Women* text sources. The upper line shows results from simulations using sub-sampled lexicons from all three texts without saccade noise. The lower line shows data based on sub-sampling from the Grimm lexicon with multiplicative noise (30% s.d.). Each data point is from a 2000-word simulation.

Table 2
Slope and intercept data for straight line fits to plots of mean saccade size vs. log lexicon size

	Brothers Grimm, no noise	*Brothers Grimm*, with noise (s.d. = 0.3)	*Glinda of Oz*	*Little Women*	542-word lexicon
Slope	−1.47	−1.04	−1.32	−1.64	−1.33
Y-intercept	15.09	10.29	14.59	15.66	13.46

Note: fits apply to all sub-lexicons shown in Fig. 2 except the smallest Grimm lexicon (10 words).

mean saccade size ranges from about 6 to 8 letters, but the dependence on lexicon size is about the same.

Table 2 lists the slopes and intercepts for straight line fits to the data sets in Fig. 2. All the slopes are close to −1 indicating that mean saccade length decreases by about one character for each log unit increase in lexicon size.

Why should this be the case? For Mr. Chips, text difficulty increases with lexicon size because uncertainty about the words increases. Appendix B describes in more detail the relationship between lexicon size, lexicon entropy, and saccade size.

Human lexicons vary widely in size, with highly educated adults typically having much larger vocabularies than children or poorly educated adults. The Mr. Chips' analysis seems to make the paradoxical prediction that mean saccade lengths are much larger in children than in the most erudite adult readers. Presumably, however, factors other than vocabulary size distinguish these groups. Mean saccade lengths are known to be shorter in children than adults, but there is also evidence that the perceptual span (similar to the visual span) is smaller in children as well (Rayner, 1986). A more direct way of assessing the lexicon-size effect in humans is to ask if the saccade size of individual readers is sensitive to the range of vocabulary in different texts. It is known that human eye movements get shorter for more difficult text (Rayner & Pollatsek, 1989, Table 4.1). Of course, range of vocabulary is only one factor contributing to text difficulty. The real question of interest is whether humans are able to activate a sub-set of their entire lexicon that is tailored to a particular context. For instance, would an article on current politics result in an effective lexicon that is different from an article on genomics? If so, we might expect contexts with a wider range of vocabulary to yield shorter saccades than contexts with smaller vocabularies, assuming syntactic and other higher-level sources of text complexity are equivalent. The notion that people have access to context-specific lexicons, i.e., implicit knowledge of variations in the probabilities of words in different contexts, is consistent with the latent-semantic analysis (LSA) model of human lexical acquisition and semantic structure (Landauer & Dumais, 1997). We emphasize, however, that Mr. Chips does not have a mechanism for detecting contexts from semantic or lexical cues. But if some external context-sensitive mechanism selectively activates a sub-set of the lexicon, reducing its effective size, the Mr. Chips' analysis shows that mean saccade length will be increased.

4. Context effect in reading

In Chips97, we pointed out that story context can be viewed as reducing the effective size of the lexicon at any point in text. We have just seen that Mr. Chips' saccades are longer for smaller lexicons. If the model were programmed so that its lexicon updated on-the-fly, reflecting story context, the model would exhibit a context effect. How would Mr. Chips' context effect compare with human context effects in reading?

Before addressing this question, we distinguish between global and local context effects. Any given book or story creates its own global context by sub-sampling its vocabulary from the entire lexicon of written English. We estimate the lexical entropy of written English to be 10.64 bits, from the word-frequency list in the British National corpus (Kilgarriff, 1997).

By comparison the entropy for *Grimms' Fairy Tales* is 8.79, *Glinda of Oz* 8.98, and *Little Women* 9.52 bits.

At any point within a given book, uncertainty about the next word is further reduced by the local context. Can you predict the next word in this…? Using a method described in Appendix C, we estimate that the average entropy of words in context in *Grimms' Fairy Tales* is 4.74 bits. As described in Appendix B, there is a tight coupling between lexicon entropy and lexicon size. Reducing the entropy from 8.79 bits to 4.74 bits is equivalent to reducing the Grimm lexicon size from its full value of 7504 words to 230 words. In short, the reduction in uncertainty associated with context is equivalent to using a sub-lexicon of about 230 words.

For the lower curve in Fig. 2, a reduction from a lexicon of 7504 to 230 words corresponds to a 25% increase in mean saccade size from 6.27 to 7.84 letters. From this result, we conclude that if Mr. Chips' lexicon updates in response to local context, he would exhibit a 25% increase in mean saccade size.

Klitz (2000) has directly compared eye movements for human participants who read simple English sentences and random text. Averaged across four participants, the mean saccade size for sentences was 32% larger than for random text. Morton (1964) compared human reading performance for zero- through eighth-order approximations to English (Miller & Selfridge, 1950). Comparing first-order (analogous to random text read by Mr. Chips) and eighth-order (very similar to English sentences), Morton found a context advantage of 33% in reading speed. Because he measured no difference in fixation times, this difference can be attributed to an increase in mean saccade length. The similarity in context advantages for humans (33%) and Mr. Chips (25%) suggests that reduced lexical ambiguity (i.e., increased predictability of words) plays an important part in explaining context effects on human reading speed.

How does Mr. Chips' mean saccade size of 7.84 characters for "in-context" reading (i.e., reading with a 230-word lexicon) compare with human values? Rayner and Pollatsek (1989, p. 118), report that people's mean saccade sizes depend on text difficulty, ranging from slightly less than 7 for science texts to slightly more than 9 for light fiction. It is unknown whether this difference is due to inherent differences in the predictability of words in these sources or to some higher-level comprehension constraint. In the Klitz (2000) study cited above, the mean saccade size for subjects reading continuous text (*Grimms' Fairy Tales*) was 6.24 letters. The reason for this low mean value is unknown, but it may be related to the unusual content and vocabulary of the fairy tales.

5. Experiment 2. Fixation locations within words: comparison with human data

Much of the research on human eye movements in reading has focused on two issues: where fixations occur within words (cf. O'Regan, 1990), and the duration of fixations (cf. Just & Carpenter, 1980; Reichle, Pollatsek, Fisher, & Rayner, 1998). Since Mr. Chips does not take time into account, only the "where" issue is relevant to the model.

For Mr. Chips, a word is said to be fixated if the central slot in the visual span falls on one of the letters of the word. Because letter information is gathered in parallel across the visual span, the central slot has no preferred status. Nevertheless, the model's "fixation" locations can be analyzed for comparison with human data.

5.1. Word skipping

We begin by asking how many words are fixated, and how many skipped. Some models of reading accord great importance to the fixated word as a marker of online cognitive processing (cf. Just & Carpenter, 1980). For Mr. Chips, words are skipped if the central slot never falls on the word. Despite "skipping" these words, Mr. Chips identifies them using letter information from other retinal slots. From a computational point of view, all words are analyzed and identified, whether they are "skipped" or "fixated."

In Chips97, we showed that Mr. Chips word-skipping behavior was strikingly similar to human word-skipping behavior for short and medium length words. For longer words, there was a discrepancy. The discrepancy may have been related to the small lexicon used by the model and the attenuated word-length distribution. Here, we return to the issue of word skipping using a larger lexicon containing a wider distribution of word lengths.

Fig. 3 shows Mr. Chips' word-skipping behavior from a 40,000 word simulation using the Grimm lexicon. The human data were taken from Rayner and McConkie (1976, Table 1) and come from 10 undergraduates who read text for comprehension.

From the figure, it is clear that Mr. Chips' word-skipping behavior is similar to human word skipping for both short and long words. For Mr. Chips, many short words can be identified without fixation because they fall into a portion of the 9-slot visual span not including the fixation slot (central slot). For example, the word "at" in the normal retina of Fig. 1 lies to the left of the central slot but is recognizable within the visual span. For longer words, it is less likely that Mr. Chips will achieve unambiguous interpretation without the word encroaching on the central slot.

Brysbaert and Vitu (1998) have divided explanations of human word skipping into two types of models: those in which saccade size is under "autonomous control" unrelated to the local lexical context, and those in which

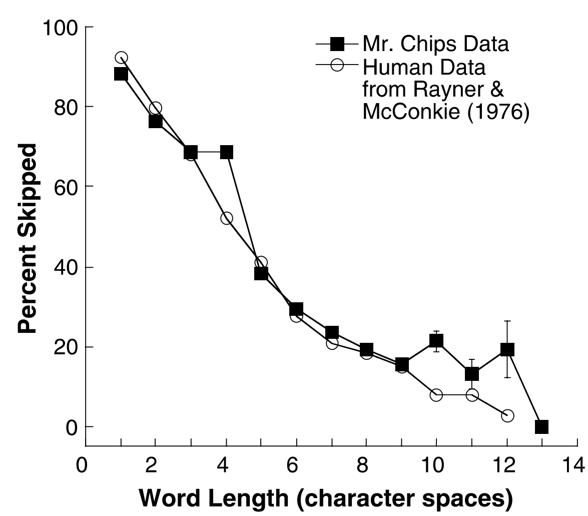

Fig. 3. Word-skipping data from human subjects and Mr. Chips. The human data are replotted from Rayner and McConkie (1976). Mr. Chips' data are based on 40,000-word simulations using the Grimm source. For Mr. Chips, a "skipped word" is defined as one for which the central slot of the model's retina never lands on the word. Error bars show the standard deviations from the binomial distribution. The error bars are larger for the long word lengths because these words were encountered only rarely in Mr. Chips' text. For word lengths from 1 to 9, the error bars are smaller than the symbols. The binomial standard deviation for word length 13 is 0 because Mr. Chips did not skip any of these words.

recognition of letters in the word to the right of fixation (termed ''parafoveal preview'') provide sufficient information for word skipping. Mr. Chips falls into the latter category; the model skips words when sufficient information about the word is available from the visual span to the right (or even left) of the nominal fixation point to unambiguously identify the word. It is reasonable to propose that a similar explanation accounts for human word-skipping behavior.

The word-skipping curve for Mr. Chips is more jagged than the human curve. The shoulder in the Chips' curve at word length 4 also showed up in simulations using lexicons from *Little Women* and *Glinda of Oz*. Jaggedness in the tail of the Chips' curve, for word lengths of 10 and greater, are probably due in part to the small number of long words encountered. (In a simulation of 40,000 words from the Grimm lexicon, there were only 31 words of length 12 and 13 words of length 13.) The fine structure of the Chips' curve may also be related to the rectangular profile of the model's visual span.

5.2. Location of fixations within words

The location of fixations within words has also been studied intensively. In two definitive papers, McConkie et al. have described the locations of initial fixations in words (McConkie, Kerr, Reddix, & Zola, 1988) and the probabilities of refixations within words (McConkie, Kerr, Reddix, Zola, & Jacobs, 1989). Their data were comprised of over 40,000 eye fixations from 66 college students who each read two chapters from a contemporary novel. The text was displayed on a computer screen, one line at a time. Eye movements were measured with an SRI dual Purkinje image eye tracker. Human data, replotted from the McConkie papers in Figs. 4–8, were digitized from the published figures using DataThief in conjunction with Kaleidagraph.

Simulation data from Mr. Chips were based on real text passages extracted from *Grimms' Fairy Tales*. As described in Section 2, we found no differences in Mr. Chips' saccade distributions between real and random text. Nevertheless, it is possible that the higher-order saccade properties determining launch/landing site statistics could be influenced by details of the sequences of words, so we used real texts for these simulations.

We will focus on two findings from McConkie et al. (1988) regarding initial fixation locations in words. First, they replicated previous research (O'Regan, 1981; Rayner, 1979) by finding that for words of length 4 letters or more, the most frequent initial fixation is at the center or just left of center of the word. Second, the more leftward the launch site of saccades, the more left the average landing site within words. We will term this the *launch-site effect*.

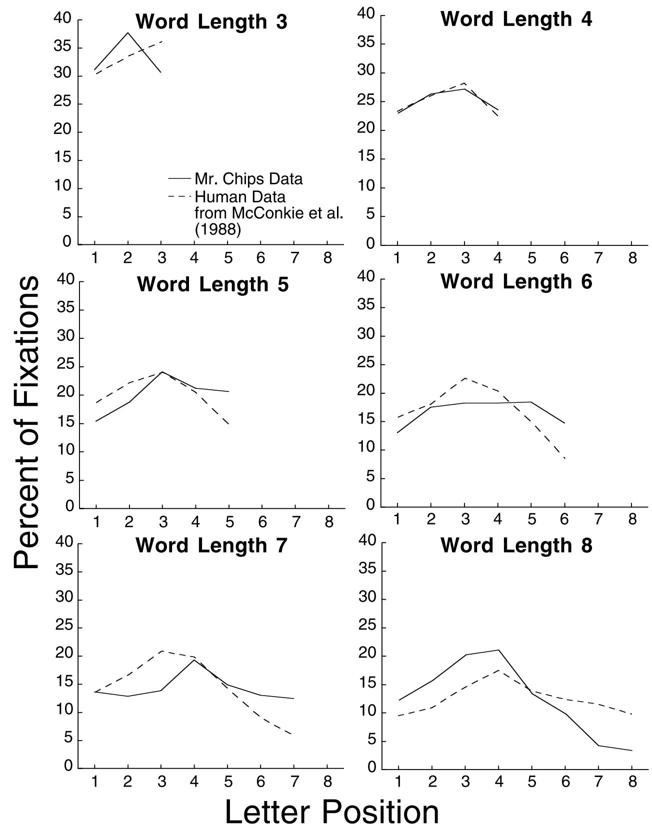

Fig. 4. Percent of first fixations as a function of letter position within words. Each panel shows data for humans and for Mr. Chips for one word length. For Mr. Chips, a fixation is defined as any time the central retinal slot lands on one of the letters in a word. The human data are replotted from McConkie et al. (1988). The Mr. Chips' data are based on 10,000-word simulations using the Grimm source, read with multiplicative saccade noise having 30% standard deviation.

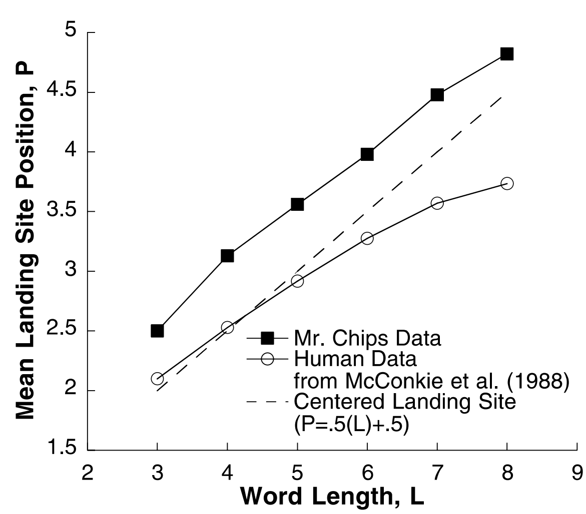

Fig. 5. Mean landing site positions, computed from the curves in Fig. 4, are plotted as a function of word length for human subjects and Mr. Chips. The dashed line shows where fixations at the middle of the word would lie.

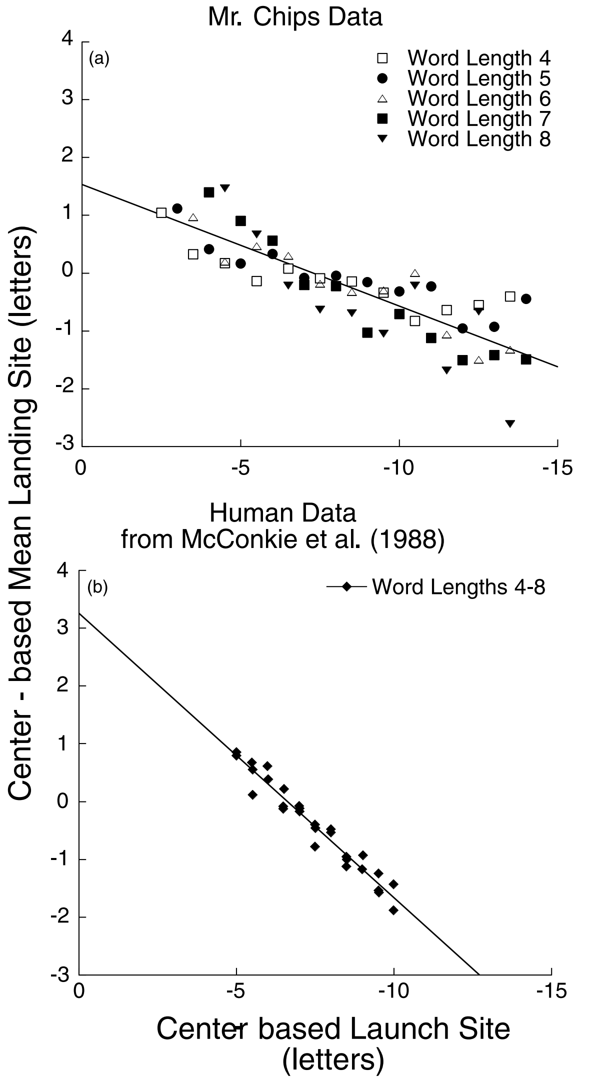

Fig. 6. Mean landing sites within words are plotted as a function of launch-site location for humans and for Mr. Chips. Both landing sites and launch sites are measured as the number of characters leftward from the center of the fixated word. Data are shown for word lengths from 4 to 8. Human data are replotted from McConkie et al. (1988). The solid lines in both panels represent best fits for all of the data shown. The Mr. Chips' best-fit line has a slope of −0.21. The human line has a slope of −0.49. Mr. Chips' data are based on 10,000-word simulations using the Grimm source, read with multiplicative saccade noise having a standard deviation of 30%.

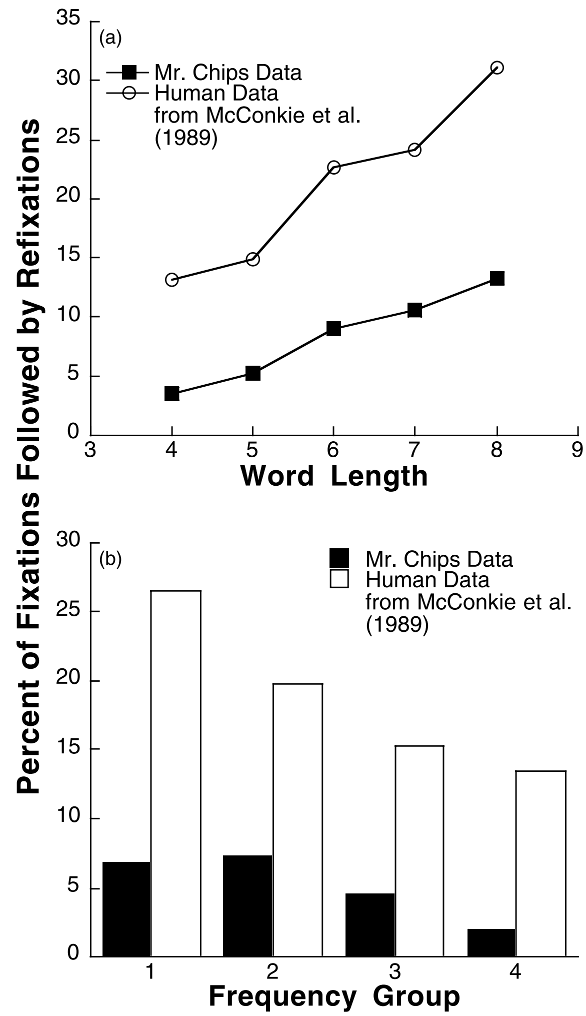

Fig. 7. Panel A shows the percentage of refixated words as a function of word length for human subjects from McConkie et al. (1989), and Mr. Chips. Panel B shows the same data, but words are grouped by the frequency with which they occur rather than their length. Frequency groups are defined by taking the common-log of the sum of one plus the numerical frequency of the word. 0–0.99 = Group 1, 1.0–1.99 = Group 2, 2.0–2.99 = Group 3, and 3.0 and above = Group 4. Mr. Chips' data are based on 40,000-word simulations using the Grimm source, read with multiplicative saccade noise having a standard deviation of 30%.

Fig. 4 displays the percentage of initial fixations [3] as a function of letter position within words. The six panels show results for words ranging in length from 3 to 8 letters. The human data are replotted from Fig. 1 of McConkie et al. (1988). Each panel shows human and Chips' data for the same word length.

The human and Chips' curves show qualitative similarity, both tending to peak near the center with broader, flatter distributions for longer words.

The graph reveals a difference in central tendency for the human and Chips' curves. The mean landing positions for Mr. Chips are consistently rightward in the word from the human values. The offsets range from 0.41 characters for 3-letter words to 1.09 character for 8-letter words, with an average offset of 0.73 characters.

This discrepancy could be resolved if the human visual span were asymmetric around the point of fixation,

[3] Skipped words are not included in this analysis. The summed percentages across letter positions in each panel add to 100%. McConkie et al. (1988) included counts for position 0, the space preceding words. In replotting their Fig. 1 data, we have omitted these fixations on spaces, and rescaled their numbers to represent the percentages of fixations on letter positions 1 through N within words of length N.

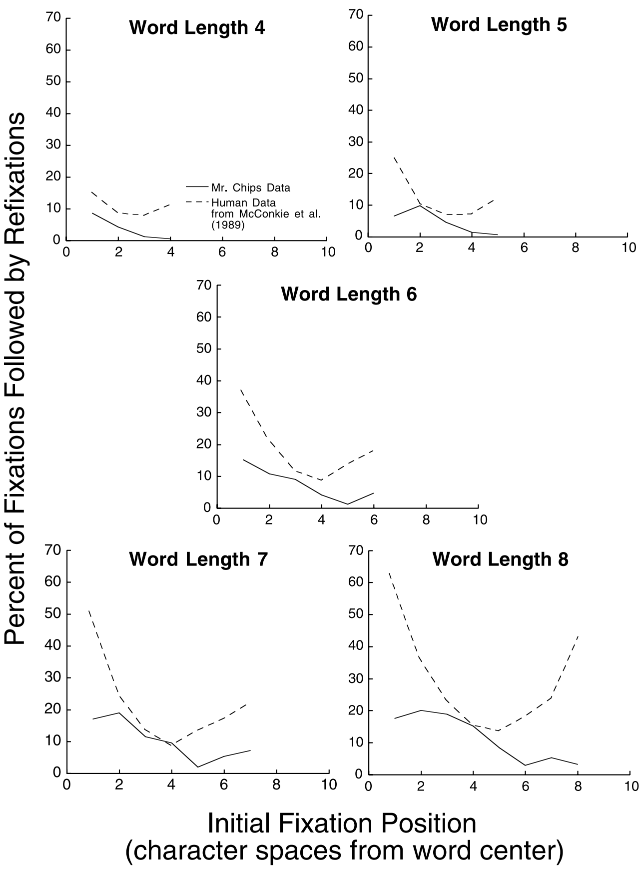

Fig. 8. Percent of refixations as a function of the position in the word of the first fixation. The human data were replotted from McConkie et al. (1989). The Mr. Chips curves are based on 40,000-word simulations from the Grimm source read with multiplicative saccade noise having a standard deviation of 30%. All of the Mr. Chips curves are based on between 109 and 179 refixations.

being slightly broader to the right than the left. If so, the geometrical center of the human visual span would be rightward of the foveal fixation point and would fall closer to the mean landing sites for Mr. Chips. Empirical measurements of visual-span profiles by Legge et al. (2001) show just this type of asymmetry. Similar asymmetries for human letter recognition have been documented by Bouma (1970) and Nazir et al. (1991).

As with word skipping, previous explanations for the locations of first fixations within words fall into two general categories. By one account, landing sites are free from linguistic influences and are determined by perceptual factors such as word length and aspects of oculomotor control (McConkie et al., 1988; O'Regan, 1990). By another view, the landing site in word $N+1$ is strongly influenced by the number of leading letters recognized in parafoveal preview of the word during the prior fixation on word N (Rayner, Sereno, & Raney, 1996). Mr. Chips encompasses both of these approaches because the model takes into account any available information about known letters, word length, and saccade noise in planning a saccade to minimize expected uncertainty about the target word.

Clark and O'Regan (1999) have presented a computational analysis of landing sites within words which is similar in spirit to ours. In their analysis, a processor has three sources of information about a word: (1) it can identify the two letters nearest fixation and the end letters of the word; (2) it knows word length, and (3) it knows the set of words in the lexicon. For any given fixation location within a word, these constraints reduce the set of possible words to a number termed the "ambiguity". For example, if fixation lies between a and t in the word *scattered*, the processor knows that a 9-letter word is present with the pattern s∗at∗ ∗ ∗∗d. If only three words in the lexicon match this pattern, the ambiguity is 3. Clark and O'Regan's ambiguity measure is analogous to the entropy measure in Mr. Chips. They showed that average ambiguity is a U-shaped curve of fixation position within words for word lengths of 5 letters and up, qualitatively similar to the human and Chips' curves in Fig. 4. Although these authors do not incorporate their ambiguity measure into a model of eye-movement control in reading, their findings are consistent with the idea that information about words is maximized (entropy or ambiguity minimized) for landing sites near the centers of words.

McConkie et al. (1988) also showed that the landing-site curves (like those in Fig. 4) are affected by the starting position of the saccade, termed launch site. For launch sites farther to the left, the landing-site distributions shift slightly leftward (McConkie et al., 1988, Table 1 & Fig. 2). In other words, the more leftward the launch site of the saccade, the greater the undershooting tendency. They further showed that the effect is due to the distance of the launch site from the center of the target word (measured in character spaces) and is independent of the length of the target word.

Fig. 6 shows summary launch-site data replotted from McConkie et al. (1988, Fig. 3), and corresponding Chips' simulation data. Each point is the average position of a landing-site distribution, plotted as a function of the launch site, both measured from the center of the target word. (Target words ranged in length from 4 to 8 characters.)

Both humans and Mr. Chips show a qualitatively similar linear relationship between launch distance and mean landing site. The effect of launch site is stronger for humans, a 0.49 character leftward shift of mean landing site for each increase of one character in launch distance, compared with only a 0.21 leftward shift for Mr. Chips.

Landing sites leftward of the center of words have sometimes been termed "undershoots", with the implication that they represent systematic errors in saccade targeting. A priori, saccade undershooting in human reading is puzzling. If undershoots have an adverse effect on performance, why do not people learn to compensate for them with longer saccades? Mr. Chips undershooting tendency is a consequence of the model's ideal computation and may explain a portion of the larger human effect. For the model, the undershoot is a consequence of accuracy considerations related to multiplicative saccade noise. Multiplicative noise means that longer saccades are associated with broader landing-site distributions. All else being equal, Mr. Chips would prefer shorter saccades because of their increased accuracy. In planning saccades toward unknown words in text, Mr. Chips implicitly computes the optimal tradeoff between positioning the visual span to reveal more of the letter information in the word, and improving accuracy by reducing saccade length. Humans may adopt an undershooting strategy to achieve a similar goal. In support, there is evidence for multiplicative saccade noise in the McConkie et al. (1988) data. From their Table 1, the breadths of the landing-site distributions grow with increasing launch distance.

Why is the human undershooting effect larger? One factor might be a human preference for the leading letters of words rather than end letters, perhaps useful for lexical look-up. Short, accurate saccades (appropriate for small launch distances) can safely target the center of words because the visual span is wide enough to encode the leading letters. For greater launch distances, however, saccade variability increases, and there is a greater risk that saccades targeted on the centers of words might go so far awry that the visual span fails to include the leading letters. A compensatory strategy would be to aim more leftward within target words for longer saccades.

Noteworthy in Fig. 6, Mr. Chips has some large launch distances (right end of the graph). Typically, Mr. Chips' longest saccades occur when the leading letters of long words are sufficient for unambiguous identification. If these letters are seen at the right end of the visual span, Mr. Chips saccades to the next word, without paying additional attention to the end of the long word. We speculate that humans are less capable of using lexical inference in this way, and, instead, they make a short saccade to look at the end of the long word. As a result, the launch distance to the next word will be reduced.

Finally, we turn to the question of refixations, words that are fixated more than once. Fig. 7 shows the probability of refixation as a function of word length and word frequency. The human data were obtained from McConkie et al. (1989). Panel A shows that both humans and Mr. Chips refixate longer words more often than shorter words. Panel B shows that when words are grouped by the frequency with which they occur rather than their length, both Mr. Chips and humans tend to refixate uncommon words more often than common words. [4] Overall, Mr. Chips has a lower rate of refixations than humans, refixating only 5% of words encountered, compared with 18.9% for humans.

Why does Mr. Chips refixate fewer words than humans, assuming both receive roughly similar visual input on each fixation? A difference in lexical inference provides at least a partial explanation. Mr. Chips often infers the identity of a word from partial information, avoiding the need for a refixation. For instance, 'disapp* * *' would be identified as "disappear" even though the three ending letters are not identifiable in peripheral vision, assuming "disappear" is the only consistent word in Mr. Chips' lexicon. Humans can do this kind of lexical inference, given enough time. But during high-speed reading, it may be faster and easier for them to refixate near the end of the word rather than search through the mental lexicon. Legge et al. (2001) have presented data and a model supporting the idea that humans make relatively little use of lexical inference in high-speed reading.

O'Regan (1984) showed that the probability of refixating a word increases when initial fixations are farther from the center of the word. His study dealt with the identification of individual words. McConkie et al. (1989) measured the probabilities of refixation for words in reading. The human data in Fig. 8 are replotted from McConkie et al. (1989, Fig. 1). As illustrated, they found that curves of refixation probability vs. initial landing site were roughly parabolic with a minimum near the center of the word. If the initial fixation occurred at the beginning or end of the word, the probability of a refixation on the same word was much greater than if the initial fixation was at the center.

McConkie et al. attributed the refixation parabolas to decreasing spatial resolution away from the point of fixation. If the end letters of a long word were fixated, then the letters at the other end would be hard to resolve, requiring a refixation. As an alternative to explaining the effect based on decreasing spatial resolution, Clark and O'Regan (1999) used their ambiguity analysis (see above). They showed that an intentional two-fixation strategy—one placed near the beginning of a word and one near the end—could be effective in minimizing ambiguity about long words.

The Mr. Chips' data in Fig. 8 were based on simulations of 40,000 words. [5] Each curve is based on

[4] Mr. Chips exhibits a reversal of this pattern, refixating Group 2 words slightly more often than Group 1 words. Additional analysis of the model indicates that this effect is robust. We believe that the effect is due to a greater number of words in Group 2 with relatively uninformative prefixes and suffixes, such as "ing".

[5] Of the 40,000 words of text, 12,260 had lengths from 4 to 8 letters. 5.7% of these words were refixated, for an average of 144 refixated words per word length.

between 109 refixations (word length 8) and 179 refixations (word length 4). The Chips' refixation data differ from the human data in several notable ways. First, consistent with Fig. 7, Mr. Chips has fewer refixations overall. Consistent with the difference in mean landing-site position (Fig. 5), the minima in the Chips' curves lie rightward of the minima in the human curves. Although both the human and Chips' curves in Fig. 8 are asymmetric, with greater probability of refixation when initially fixating leading rather than ending letters, the asymmetry is much greater for Mr. Chips.

Why does the optimal performance of Mr. Chips result in refixation curves that are so asymmetric, with low probability of refixation when the first fixation is at the end of a word? Mr. Chips will make an initial fixation near the end of a long word when the leading letters of the word are encoded on a prior fixation. For example, suppose the rightmost four slots in his visual span and his right periphery reveal the pattern 'rece###_', that is, a 7-letter word beginning 'rece'. The word could be "receive" or "receipt" (assuming these words are in his lexicon). Since none of the visible letters falls in the central slot of his 9-slot visual span, he has not yet "fixated" this word. Nevertheless, an optimal saccade for resolving uncertainty is to look at the end of the word. The resulting noisy saccade could end up with the sixth or seventh letter of the word falling in the central slot of the visual span. This initial fixation on the word would resolve the uncertainty.

Why do not humans behave like Mr. Chips in this respect? The greater symmetry of the human refixation "parabolas" may imply less benefit from letter recognition on prior fixations (except perhaps when the entire next word is legible on prior fixations—see above on word skipping). Perhaps humans are less adept at integrating letter information across inter-word saccades, or perhaps they are locked into a centering saccade strategy (O'Regan, 1990) even when targeting of the word ending would be optimal for word recognition. Initial fixations at the ends of words by humans may occur only as a result of saccade error, not by intention as with Mr. Chips.

The differences in refixation behavior between Mr. Chips and humans (Fig. 8) may be put into perspective as follows. Out of every 100 words fixated, humans refixate on about 19 words and Mr. Chips on about five. Suppose the extra 14 fixations made by people are due to inferior lexical inference (i.e., they make an eye movement to an unrecognizable part of the word because it is faster or easier to do so than to infer the word's identity from lexical inference), and assume that the remaining saccades are all optimal like Mr. Chips. Then, roughly speaking, humans will make about 14% more saccades than Mr. Chips, and mean saccade length will be about 14% smaller (say, about 6 letters compared to 7). This is about the size of the difference we found in Chips97 between the ideal observer and a sub-optimal model who made no use of lexical inference. In terms of mean saccade size, the extra refixations by humans (possibly due to failures of lexical inference) may have a relatively modest impact. But, when we isolate the words in which refixations occur, differences in the capacity for lexical inference between humans and the ideal observer are magnified.

6. Summary and conclusions

In this paper, we have compared the behavior of an ideal-observer model of reading to data in the literature on normal human reading. Simulations were conducted with parameters of the Mr. Chips' model set to emulate human values—a visual span of 9, and multiplicative saccade noise with a standard deviation of 30%. We now summarize the major findings.

6.1. Lexical analysis

The analyses in Chips97 were limited to results from a small lexicon with relatively few long words. Here, we have reported results for lexicons ranging in size up to 11,028 words from three full-length children's books in addition to the original 542-word Chips' lexicon. The major results were:

(1) Mean saccade length decreases by about one character for each log unit increase in lexicon size. This finding is equivalent to the rule that mean saccade size decreases by one character for each 2.5 bits increase in lexicon entropy (Appendix B).

(2) If Mr. Chips' lexicon were updated in response to local context, his mean saccade size would increase by 25%. Humans exhibit an increase of about 30% in mean saccade length for sentences compared with random strings of words. Accordingly, Mr. Chips' behavior provides an explanation of context effects on human reading speed in terms of the predictability of words in text (entropy reduction).

6.2. Fixation locations within words

We compared Mr. Chips' reading behavior to humans in word skipping, the location of first fixations within words, and the probability of refixations within words.

(1) Mr. Chips "skips" a word if the central slot of his visual span never lands on the word. The "skipped" word is identified from information gathered elsewhere in his visual field. The dependence of word skipping on word length is very similar for Mr. Chips and humans, suggesting that the same front-end factors explain human word skipping.

(2) Like humans, Mr. Chips' initial fixations tend to lie near the center of words. But, there is a systematic

difference; on average, Mr. Chips' mean landing position is 0.73 characters more to the right than humans. This difference can be explained if the human visual span is asymmetric, being slightly broader to the right of the point of fixation than to the left. Human data exhibit such an asymmetry.

(3) The location of initial fixations within words is affected by the launch distance, i.e., the length of the just-completed saccade. The effect is stronger for humans, a 0.49 character shift leftward in mean landing site for each increase of one character in launch distance, compared with only a 0.21 shift for Mr. Chips.

(4) Mr. Chips has a much lower rate of multiple fixations on the same word than humans, 5% vs. 18.9%. Mr. Chips' lower rate of multiple fixations is attributable to the superior ability to use lexical inference to identify words from partial information.

(5) Humans are more likely to refixate words when their initial fixations fall at the beginning or the end of the word. Different from humans, Mr. Chips rarely refixates a word when his initial fixation is at the end of the word. This difference can be understood if initial fixations at the ends of words by humans occur only as the result of saccadic error, not by strategic intention.

6.3. Use of lexical knowledge

Both of the experiments described in this paper dealt with lexical knowledge, but in different ways. Experiment 1 was concerned with the impact on saccades of the size of the currently active lexicon. Mr. Chips' saccades tend to be longer for smaller lexicons, because there is less ambiguity in word identification. We argued that a human context effect—longer saccades for real vs. random text—might be explained as the online activation of appropriate sub-lexicons. In Experiment 2, we discussed a second process, termed lexical inference. Given any particular lexicon, Mr. Chips often infers word identity from partial information, e.g., disap∗ ∗ ∗∗ is "disappear" if there is no other consistent word in the lexicon. If humans can do this kind of lexical inference at all, they may be slow, and it may be faster for them to make an additional saccade to the unseen end of the word. Although Mr. Chips is optimal at both types of lexical processes, our analysis suggests that humans are good at doing the first of these, but not the second.

Mr. Chips uses an optimizing principle (entropy minimization) to plan saccades. Humans may not do this precise computation, but they appear to adopt heuristics that approximate the outcome of this calculation, in the sense that a median approximates the computation of a mean value. These heuristics may involve eye-movement targeting strategies similar to those described in the strategy tactics model (O'Regan, 1990), and may also include on-the-fly adjustments in the effective lexicon. Where humans may depart substantially from true entropy minimization is in intra-word lexical inference where it may be easier and faster to make refixations than to rely on lexical analysis to reduce ambiguity. There is no contradiction in saying that the human algorithm is equivalent to the ideal observer Mr. Chips for some aspects of its performance, and not others. In principle, knowing when the human algorithm departs from the ideal observer provides information about the algorithm.

The Mr. Chips' model provides a method for assessing the impact of visual, oculomotor and lexical constraints on the complex task of reading. When human reading behavior parallels the model's behavior, as in numerous cases discussed in this paper, the model provides an account of human performance. As we have also seen, when the model differs from humans, the task of accounting for these differences can provide further insight into human reading.

Acknowledgements

We thank Charles R. Fletcher and Paul van den Broek for helpful discussion of context effects in reading. We also thank two anonymous reviewers for their detailed and insightful commentaries. This research was supported by NIH grant EY02934 to Gordon E. Legge.

Appendix A. Example of Mr. Chips' entropy-minimization algorithm

In Fig. 9, Mr. Chips has a visual span of size one, that is, one high-resolution slot and four low-resolution peripheral slots. Here, Mr. Chips knows only that we have a four-letter word that begins with the letter 'c'. An examination of his 542-word lexicon indicates that there are five words that fit this pattern: "call", "came", "city", "cold", and "come". The conditional probabilities in Fig. 9 are the probabilities that 'c∗ ∗ ∗' is each of the five possible words, based on word frequencies kept in the lexicon. Mr. Chips is goal is to make a saccade to minimize the uncertainty of the current word. From information theory, we define uncertainty as the entropy H:

$$H = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

where p_i is the conditional probability of each possible word, and k is the number of possible words. In this case, the entropy is 2.10 bits of information (the entropy of five equally likely candidate words is 2.31 bits).

In order to determine which saccade size, if executed, minimizes the entropy, Mr. Chips must consider all possible saccade sizes. In this example (see Fig. 9(B)),

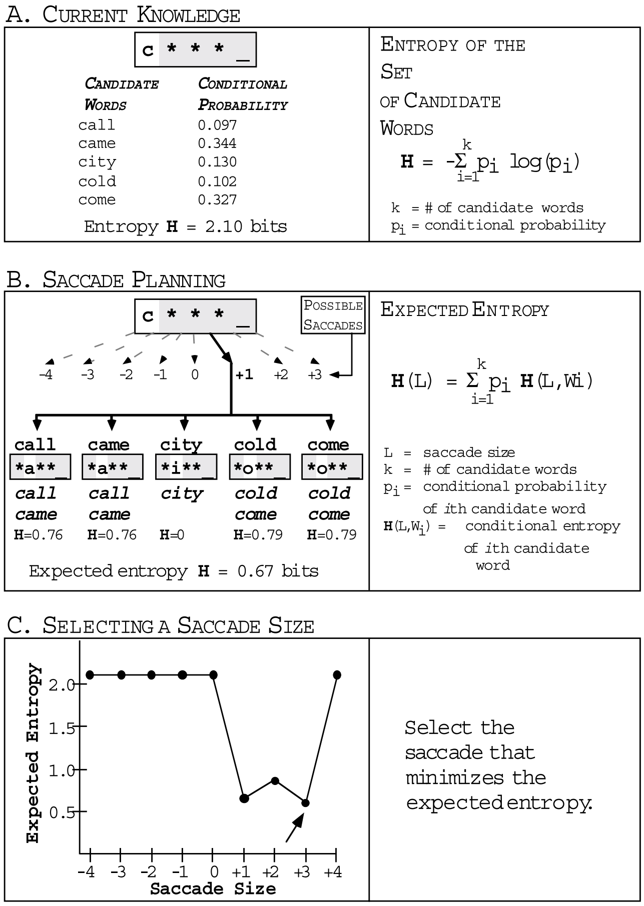

Fig. 9. Illustration of Mr. Chips' entropy-minimization algorithm. Based on his current knowledge of an unidentified word, Mr. Chips calculates the entropy reduction associated with each possible saccade and chooses the one that minimizes expected uncertainty about the word.

the range of saccades that keep the current word in the visual range are from -4 to $+3$. For each possible saccade size, Mr. Chips computes the expected entropy by considering all possible outcomes. For instance a saccade of $+1$ will reveal the second letter of the word. As Fig. 9 shows, if the word is "call", the letter 'a' will be revealed. But Mr. Chips will remain unsure whether the word is "call" or "came", with corresponding entropy of 0.76 bits. The same is true if the word is "came". If the word is "city", then the letter 'i' is revealed. Since there is only one four-letter word in the small lexicon beginning with 'ci', the uncertainty would be reduced to 0 bits. Similar calculations can be made for the outcomes of "cold" and "come" ($H = 0.79$). The overall expected entropy of a saccade of size $+1$ is the frequency-weighted average of all five possibilities. Here, the expected entropy $H = 0.67$.

We can do this same computation for all possible saccade lengths in the given range (see Fig. 9(C)). Then Mr. Chips chooses the saccade size that has the lowest expected entropy, in this case a saccade of size $+3$. This entropy analysis continues until the current word is identified, then Mr. Chips moves on to the next word.

Appendix B. Mean saccade size and lexicon entropy

From a computational perspective, lexicon entropy is a more direct determinant of Mr. Chips' mean saccade size than lexicon size per se. In this appendix, we show (1) how lexicon entropy depends on lexicon size, and (2) how Mr. Chips' mean saccade size depends on lexicon entropy.

Lexicon entropy is expected to grow with lexicon size. This relationship is shown in Fig. 10 for the Grimm sub-lexicons.

A straight line with slope 2.67 provides a good fit between entropy and log lexicon size. The dashed line above the data points has a slope of 3.3 and represents the relationship that would be found if the distribution of word frequencies within the text were uniform. [6] The other text sources showed similar curves, although the slope for *Glinda* was a little steeper. This suggests that it has a more uniform distribution of words, probably a reflection of the smaller overall vocabulary in the *Glinda* text.

Fig. 11 shows the dependence of mean saccade size on lexicon entropy for the Grimm texts. For lexicon entropies above 4 bits (corresponding to lexicon sizes of 90 words and above), a straight line provides a reasonable summary of the relationship between mean saccade size and entropy. The slope of the best-fitting straight line is 0.42 letters per bit. In other words, Mr. Chips loses roughly one letter in mean saccade length for each 2.5 bits growth in lexicon entropy. Similar relationships hold for the *Glinda of Oz* and *Little Women* lexicons.

Appendix C. Estimating the mean entropy of words in context

Here, we show how we estimated the average entropy of words in context.

We assume that most of the entropy reduction attributed to context is associated with the structure and content of individual sentences.

For individual sentences, the entropy of the first word is relatively high (It is not much easier to guess the first word of a sentence than an arbitrary word in a text.), but the entropy of succeeding words decreases as the reader proceeds through the sentence. Our method averages across these within-sentence effects.

[6] The slope of 3.3 for a uniform frequency distribution arises because 10 equally likely alternatives has 3.3 bits of entropy, 100 equally likely alternatives 6.6 bits of entropy, etc.

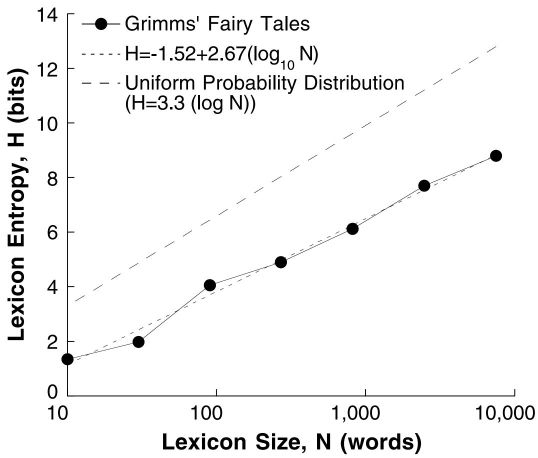

Fig. 10. Lexicon entropy is shown as a function of the base-ten log of lexicon size. Six sub-sampled lexicons were derived from the original 7504-word Grimm lexicon and their entropies were calculated (along with the entropy of the complete Grimm lexicon). The lower dashed line is a best-fit curve showing the relationship between lexicon entropy and log lexicon size. The slope is 2.67 bits per log unit of lexicon size. The upper dashed line has a slope of 3.3 and represents the curve that would result from a flat word-frequency distribution in each sub-sampled lexicon.

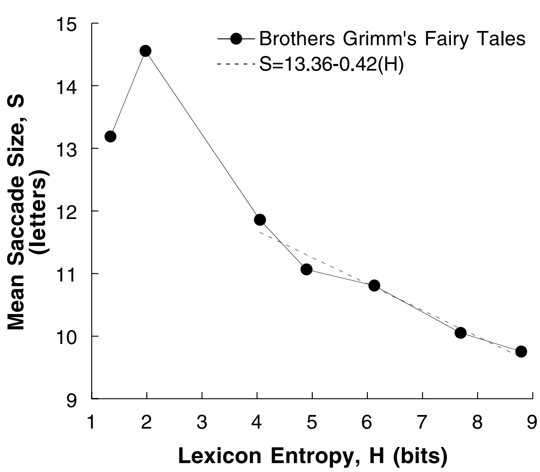

Fig. 11. Mr. Chips' mean saccade size is plotted as a function of the entropy of lexicons of different size, sub-sampled from the Grimm lexicon. Each data point is based on a 500-word no noise simulation. A best-fitting straight line is shown for the five largest lexicons with a slope of −0.429 characters per bit.

Shannon (1951) provided estimates of the entropy of an Nth letter following $N-1$ preceding letters of context. The entropy drops as N increases. We fit the following curve to Shannon's lower bound entropy estimates:

$$H(N) = 2.943 \times N^{-0.3774}$$

where H is the entropy of the Nth letter in context.

Assuming an average sentence is 60-characters long, we can use this formula to estimate the entropy of each of the 60-character positions. Summing across these 60 values, we find the total sentence entropy to be 57.3 bits.

In *Grimms' Fairy Tales*, the average word length is 3.94 letters.

Taking spaces into account, a 60-character sentence would have an average of 12.15 words. Dividing the total sentence entropy of 57.3 bits among these 12.15 words gives an average entropy per word of 4.74 bits.

References

Alcott, L. M. (1869). *Little Women*. Available: http://www.gutenberg.net.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: activation or search? *Journal of Experimental Psychology: Learning Memory, and Cognition, 15*, 802–814.

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.

Baum, L. F. (1920). *Glinda of Oz*. Available: http://www.gutenberg.net.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177–178.

Brysbaert, M., & Vitu, F. (1998). Word skipping: implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception*. Oxford: Elsevier.

Burgess, A., & Barlow, H. B. (1983). The precision of numerosity discrimination in arrays of random dots. *Vision Research, 23*, 811–820.

Carroll, J. B., Davies, P., & Richman, B. (1971). *The American heritage word frequency book*. Boston: Houghton Mifflin.

Chung, S. T. L., Legge, G. E., & Tjan, B. S. (in press). Spatial frequency characteristics of letter identification in central and peripheral vision. *Vision Research*.

Clark, J. J., & O'Regan, J. K. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research, 39*, 843–857.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review, 96*, 267–314.

Grimm, J., & Grimm, W. (1812). *Grimm's Fairy Tales*. Available: http://www.cs.cmu.edu/~spok/grimmtmp.

Hecht, S., Shlaer, S., & Pirenne, M. H. (1942). Energy, quanta, and vision. *Journal of General Physiology, 25*, 819–840.

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology, 26*, 240–346.

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

Kilgarriff, A. (1997). Putting frequencies in the dictionary. *International Journal of Lexicography, 10*, 135–155.

Klitz, T. S. (2000). *The effect of visual span size and simulated central scotomas on reading performance*. Doctoral Dissertation. University of Minnesota.

Klitz, T. S., Legge, G. E., & Tjan, B. S. (2000). Saccade planning in reading with central scotomas: comparison of human and ideal performance. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process*. New York: Elsevier.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*, 211–240.

Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading. XVI. The visual span in normal and low vision. *Vision Research, 37*, 1999–2010.

Legge, G. E., Gu, Y., & Luebker, A. (1989). Efficiency of graphical perception. *Perception and Psychophysics, 46*, 365–374.

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: an ideal-observer model of reading. *Psychological Review, 104*, 524–553.

Legge, G. E., Mansfield, J. S., & Chung, S. T. L. (2001). Psychophysics of reading. XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research, 41*, 725–734.

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research, 35*, 549–568.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

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., Kerr, P. W., Reddix, M. D., Zola, D., & Jacobs, A. M. (1989). Eye movement control during reading: II. Frequency of refixating a word. *Perception and Psychophysics, 46*, 245–253.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics, 16*, 578–586.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and recall of meaningful material. *American Journal of Psychology, 63*, 176–185.

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.

Nazir, T. A., O'Regan, J. K., & Jacobs, A. M. (1991). On words and their letters. *Bulletin of the Psychonomic Society, 29*, 171–174.

O'Regan, J. K. (1981). The convenient viewing position hypothesis. In D. F. Fisher, R. W. Monty, & J. W. Senders (Eds.), *Eye Movements: Cognition and Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Regan, J. K. (1984). How the eye scans isolated words. In A. G. Gale, & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research*. Amsterdam: North-Holland.

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 and Psychophysics, 34*, 457–464.

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31*, 1399–1415.

Pelli, D. G., Burns, C. W., Farell, B., & Moore, D. C. (in press). Identifying letters. *Vision Research*.

Pollatsek, A., Binder, K. S., & Perea, M. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1142–1158.

Rayner, K. (1979). Eye guidance in reading: fixation locations within words. *Perception, 8*, 21–30.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41*, 211–236.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research, 16*, 829–837.

Rayner, K., & Pollatsek, K. (1989). *The psychology of reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1188–1200.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125–157.

Reilly, R. G., & O'Regan, J. K. (1998). Eye movement control during reading: a simulation of some word-targeting strategies. *Vision Research, 38*, 303–317.

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal, 30*, 50–64.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*, 395–397.

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.