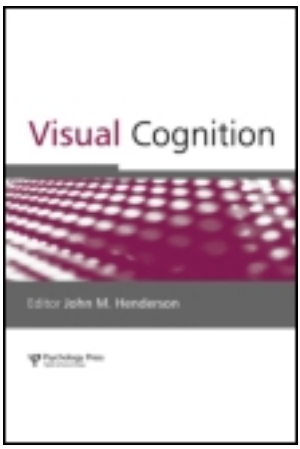

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pvis20

The utility of modelling word identification from visual input within models of eye movements in reading

Klinton Bicknell [a] & Roger Levy [b]
[a] Department of Psychology , University of California , San Diego , CA , USA
[b] Department of Linguistics , University of California , San Diego , CA , USA
Published online: 23 May 2012.

To cite this article: Klinton Bicknell & Roger Levy (2012) The utility of modelling word identification from visual input within models of eye movements in reading, Visual Cognition, 20:4-5, 422-456, DOI: 10.1080/13506285.2012.668144

To link to this article: http://dx.doi.org/10.1080/13506285.2012.668144

VISUAL COGNITION, 2012, 20 (4–5), 422–456

Psychology Press
Taylor & Francis Group

The utility of modelling word identification from visual input within models of eye movements in reading

Klinton Bicknell[1] and Roger Levy[2]

[1]Department of Psychology, University of California, San Diego, CA, USA
[2]Department of Linguistics, University of California, San Diego, CA, USA

Decades of empirical work have shown that a range of eye movement phenomena in reading are sensitive to the details of the process of word identification. Despite this, major models of eye movement control in reading do not explicitly model word identification from visual input. This paper presents an argument for developing models of eye movements that do include detailed models of word identification. Specifically, we argue that insights into eye movement behaviour can be gained by understanding which phenomena naturally arise from an account in which the eyes move for efficient word identification, and that one important use of such models is to test which eye movement phenomena can be understood this way. As an extended case study, we present evidence from an extension of a previous model of eye movement control in reading that does explicitly model word identification from visual input, Mr. Chips (Legge, Klitz, & Tjan, 1997), to test two proposals for the effect of using linguistic context on reading efficiency.

Keywords: Computational modeling; Eye movements in reading; Visual word identification.

Please address all correspondence to Klinton Bicknell, Department of Psychology, University of California San Diego, 9500 Gilman Drive #109, La Jolla, CA 92093-0109, USA. E-mail: kbicknell@ucsd.edu

We are grateful to Gordon Legge for sharing the corpus used in the original Mr. Chips experiments. Portions of this work were presented at the 32nd annual conference of the Cognitive Science Society and the 84th annual meeting of the Linguistic Society of America. The research was supported by NIH training grant T32-DC000041 from the Center for Research in Language at UC San Diego to KB and by a research grant from the UC San Diego Academic Senate, NSF grant 0953870, and NIH grant R01-HD065829, all to RL.

http://www.psypress.com/viscog http://dx.doi.org/10.1080/13506285.2012.668144

One of the major drivers of eye movements in reading is the identification of the words in the text being read (Rayner, 1998, 2009). Over the past decades, the word identification process has been demonstrated to be sensitive to the precise visual input about the word that the eyes receive: e.g., identification is most efficient when the eyes fixate the part of a particular word that provides the most disambiguating information to distinguish that word from the other words in the lexicon (e.g., Clark & O'Regan, 1999; O'Regan, Lévy-Schoen, Pynte, & Brugaillère, 1984). Given this, it is perhaps a surprising state of affairs that major contemporary models of eye movement control in reading (e.g., Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Warren, & McConnell, 2009) do not explicitly model word identification from visual input. Here, we argue that the inclusion of detailed models of word identification within models of eye movement control in reading—that is, modelling the dependence of the process of word recognition on the precise visual input the eyes receive given their position and the text around them—is a necessary and useful step towards achieving a fuller understanding of eye movements in reading.

The structure of the paper is as follows. First, we review empirical evidence that many aspects of eye movements in reading can be understood as naturally arising from an account in which the eyes are efficiently directed to identify words from visual input, which we will refer to as an *efficient visual identification account*. We next describe how the major models of eye movement control in reading (Engbert et al., 2002, 2005; Reichle et al., 1998, 2009) can account for many (but not all) of these effects without incorporating detailed models of word identification from visual input, and make an argument for why developing models that do incorporate detailed models of word identification from visual input can be useful: To verify proposals for how eye movement phenomena can be understood as arising from efficient visual identification. We then describe the first model of eye movements in reading that meets this criterion, Mr. Chips (Legge, Hooven, Klitz, Mansfield, & Tjan, 2002; Legge, Klitz, & Tjan, 1997), and show how it verifies that a number of such proposals can in fact produce the eye movement behaviour in question. Finally, as an extended case study on the utility of models that include models of word identification from visual input, we focus the remainder of the paper on using such models to test two intuitions for the effect of linguistic context.

READING ISOLATED WORDS

The first suggestion that the word identification process is sensitive to the precise visual input that the eyes receive about the word was given by Rayner

(1979). Rayner reported evidence that the median landing site of the eyes on a word of text in natural reading of English is just left of its centre, and conjectured that one explanation for this *preferred viewing location* is that it may provide the maximum information about the word being fixated. This hypothesis was further elaborated by O'Regan (1981), who suggested that the most efficient place to look in a word is the one that will provide the most disambiguating visual information to distinguish the word from its visual neighbours—what is now referred to as the *optimal viewing position*.[1] O'Regan's (1981) suggestion implies that, although this position may be on average just left of centre, it should be different for each particular word, since each word has a different distribution of visual neighbours. For example, the word *xylophone* has a very rare beginning, and knowing just the first two letters distinguish it from virtually all other words of English; by contrast, knowing that the first two letters of a word are, e.g., "ca" leaves a large range of possible word identities. Thus, O'Regan predicts that the optimal viewing position should differ between words.

A range of studies of isolated word recognition have supported the basic hypothesis that words are on average identified most quickly when fixated at or just left of the word centre. In the studies, the word to be identified is displayed at varying displacements relative to the point of fixation, allowing for experimental control over the position in the word that the reader is fixating. Then, the researcher can examine the time it takes to say the word aloud (naming latency), the probability of making a second fixation on the word (refixation probability), or the total duration of all fixations on the word (gaze duration), each as a function of the reader's initial fixation position within the word. Using this methodology, O'Regan et al. (1984) presented evidence that—when aggregating over a range of words—naming latencies, refixation probabilities, and gaze durations were all U-shaped functions with a minimum at a point just left of the word centre, which climb rather sharply as the fixation is further away from this point, supporting the notion that the optimal viewing position is on average just left of a word's centre. The conclusion that this effect arises because this is the location in the word that provides the most disambiguating information to distinguish a word from its neighbours is further supported by lexical analysis conducted by Clark and O'Regan (1999), which demonstrated formally that—under certain assumptions about the nature of the visual input obtained from a word given the point of fixation—ambiguity about the identity of a word is minimized on average by fixating just left of centre, vindicating O'Regan's original argument.

[1] O'Regan originally referred to this position as the *convenient viewing position*.

Another result from these studies concerns the locations in the word to which refixations take the eyes. We can contrast two hypotheses about why a reader would make a refixation to efficiently identify a word. The hypothesis that word identification involves obtaining enough visual information to disambiguate a word from its visual neighbours predicts that if enough visual information to disambiguate the word cannot be obtained from the current fixation location, then a reader must move their eyes to another part of the word—a natural motivation for refixations. Conversely, an alternative explanation for the optimal viewing position phenomenon may be that word identification efficiency is related to how well the word can be seen overall (e.g., perhaps the average distance of the eyes from each of the word's characters). Under this hypothesis, if a reader's initial fixation location is sufficiently far from the optimal viewing position, the most efficient correction would be to move the eyes to the optimal position. Thus, when the initial fixation is too close to the beginning of the word, the disambiguation hypothesis would predict that a refixation should take the eyes to the end of the word, whereas the average distance hypothesis would predict that a refixation should take the eyes to the optimal position (just left of centre). O'Regan and Lévy-Schoen (1987) report that refixations in these single-word identification tasks typically take the eyes to the other side of the word, thus supporting the hypothesis that the best way to understand these effects is that obtaining disambiguating visual input is crucial to how readers identify words.

Finally, data from these experiments also support O'Regan's (1981) proposal that the optimal viewing position should vary across words, depending on where the most useful disambiguating visual information is located for a particular word relative to its visual neighbours. In experiments comparing French words that could be uniquely disambiguated given only information about the end of the word but not with information about the beginning (e.g., *circonspecte*, *interrogatif*, *transversal*, *approfondi*, *architecte*) with words that could be uniquely determined from information about the beginning but not with information about the end (e.g., *perquisition*, *attroupement*, *arrestation*, *auxiliaire*, *hirondelle*), O'Regan and colleagues showed that the optimal viewing position (defined for this task as the fixation point that minimized gaze duration) was closer to the end of the word for the former group and closer to the beginning of the word for the latter group (Holmes & O'Regan, 1987; O'Regan & Lévy-Schoen, 1987; O'Regan et al., 1984). Analogous results have also been shown for single word recognition in Finnish (Hyönä, Niemi, & Underwood, 1989). Taken together, all these results provide substantial evidence that the word identification process is one of

obtaining the specific visual information which will disambiguate a word from its visual neighbours.

READING WORDS IN TEXT

Given that one of the major components of reading continuous text is word identification, we might expect the insights obtained from the study of isolated word identification to directly transfer to reading words embedded in text. However, in reading continuous text, readers have access to two sources of information about a word before their first fixation on the word, which makes the insights gained from the isolated word recognition case transfer less directly. One additional source of information that readers can have when reading words embedded in text is visual information about the word, obtained parafoveally. For example, there is ample evidence that readers can detect the length of a word in the parafovea (e.g., Morris, Rayner, & Pollatsek, 1990; Pollatsek & Rayner, 1982; Rayner, 1979) and, in addition, can often obtain information about (or even identify) the first letters in the word following the one that is fixated (e.g., Rayner, McConkie, & Zola, 1980, Rayner, Well, Pollatsek, & Bertera, 1982). The second additional source of information about the identity of a word that readers have in advance is linguistic context. Linguistic context can provide substantial constraint on the possible words that will occur in a particular position, and has robust effects on eye movements in reading (Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981). For example, in the context "The children went outside to . . .", only the identities of the first couple of letters would be necessary to be virtually certain that the next word is *play*. Given access to these additional sources of information prior to their first fixation on a word, the point in the word that will provide the most disambiguating visual information about its identity will change.

Because of these complications, we might expect that the relationship between a measure like gaze duration and the eyes' first landing position on a word will be less direct when reading words embedded in text than words in isolation, which is precisely what Vitu, O'Regan, and Mittau (1990) found: Gaze durations are on average a relatively constant function of initial landing positions on the words. Despite the lack of direct effects on gaze durations, there are other indications that word identification in reading works in a similar fashion to isolated word recognition. For one, the familiar U-shaped curve, with a minimum just left of word centre, does appear when analysing refixation rates by initial landing position (McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; Rayner, Sereno, & Raney,

1996). Relatedly, as already mentioned, readers direct their saccades to words such that their initial fixations on them are on average at or just left of word centre. The strongest evidence, however – as with single word identification studies – is provided by demonstrations that eye movement behaviour in reading is sensitive to the location of the most useful disambiguating information in particular words. A range of sentence reading studies in English, French, and Finnish have compared eye movement behaviour when reading words that can be disambiguated given only information about the beginning of the word (i.e., words with redundant endings) to words that cannot be uniquely identified given only information about the beginning. The results show that when the word's ending is redundant, readers are more likely to skip the word's second half, and when they do fixate the second half, they do so for a shorter duration (Hyönä, 1995; Hyönä et al., 1989; Pynte, Kennedy, & Murray, 1991; Rayner & Morris, 1992; Underwood, Bloomfield, & Clews, 1988; Underwood, Clews, & Everatt, 1990). The natural interpretation of such findings is that because readers are more likely to be able to identify words with redundant endings given only visual information about their beginning, they are less likely to require further visual information about their end. Results such as these—as well as general principles of theoretical parsimony—suggest that the underlying principles of the identification of words embedded in text should be the same as for words in isolation, i.e., that identification is a process of obtaining the specific visual information that will disambiguate a word from its visual neighbours.

This emphasis on word identification from visual input also has the advantage of providing a unified explanation for a number of phenomena seen *only* in continuous reading. One example is the fact that the mode of the distribution of initial landing positions on a word shifts depending on the launch site of the previous saccade. That is, when the saccade originates from further back, the eyes tend to land closer to the beginning of the word, and when the saccade originates from a position closer to the word, the eyes tend to land closer to its end (McConkie, Kerr, Reddix, & Zola, 1988). Because fixation positions closer to the word can yield more parafoveal preview of the word's initial letters (Rayner et al., 1980, 1982), it seems reasonable to suppose that the position in the word containing the as yet unobtained visual information most useful for disambiguation may also shift to the right. Another case for which this account may provide a simple explanation is the effect of word length on the probability of skipping a word, i.e., not making a fixation on the word. Rayner and McConkie (1976) demonstrated that skipping rates decrease with word length, ranging from average rates over 90% for one-character words to around 40% for five-character words and down to 10% for 10-character words. The nature of this relationship is precisely what one

might expect if word identification occurs by obtaining disambiguating visual input. If we make the simplifying assumption that readers get approximately the same amount of parafoveal information about words of each length, for example, we might expect the probability of being able to effectively disambiguate a word based on parafoveal information to decrease as words become longer. In summary, not only are there empirical and theoretical reasons to believe that this account of word identification plays a large role in shaping eye movements in reading, but the account can also be used to provide intuitive explanations for a range of eye movement phenomena.

MAJOR MODELS OF EYE MOVEMENT CONTROL IN READING

Given the discussion in the previous two sections, it may at first seem surprising that the major models of eye movement control in reading do not model the process of word identification from visual input. Instead, models such as E-Z Reader (Pollatsek, Reichle, & Rayner, 2006; Reichle et al., 1998) and SWIFT (Engbert et al., 2002, 2005; Schad & Engbert, this issue 2012) make a small set of assumptions about how the word identification process works on average, and define the model to behave relatively reasonably given those assumptions. Specifically, the single assumption made by these models about the word identification process that relates to visual input[2] is essentially that the rate of word identification is lower for words that are further away from the point of fixation. Formally, in E-Z Reader, the rate of word identification decreases with the average distance of each letter in the word from the point of fixation. In SWIFT, the situation is slightly more complex. There, a processing rate function assigns each letter position a processing rate. These rates decrease with distance from the point of fixation, but fall off more rapidly on the left than the right. Then, the rates of word identification in the model are defined to be faster as the mean of the processing rates of the letters comprising the word increases (similar to E-Z Reader) and also as the sum of the rates of the letters increases. Note that these simplified models of word identification abstract over any properties of particular words, and there is no representation of how the letter or word in question is distinguished from the other possible letters or words that might have been present.

Despite these simplifications, it is the case that under each of these formalizations, the optimal viewing position (i.e., the position from which a word will be identified most quickly) will be near the empirically

[2] There are also other assumptions made about the word identification process that do not relate to visual input, which encode effects of word frequency and predictability.

determined optimal viewing position. For E-Z Reader, it is apparent that it will be exactly at the centre of the word, since the word identification rate decreases with distance from the word centre. Similarly, for SWIFT, it will also be near the centre of the word, since the word identification rate decreases as the mean and sum of the processing rates of the letters decrease. For SWIFT, however, because the visual acuity function is asymmetric and falls off more sharply to the left than to the right, these rates will be highest when fixating slightly left of the word centre. Thus, this single assumption correctly encodes the fact that, on average, words in isolation will be recognized most quickly when fixated near the centre, a fact that would result naturally from the inclusion of a model of word identification from visual input.

Given this assumption about the word identification process, one reasonable way for the models to behave would be to always target saccades to the centre of words (where they can be most efficiently identified), and this is in fact what both models do. This stipulation thus encodes the fact that initial fixations on words are near the word centre, without appealing to any details of the word identification process. Additionally, both models assume that there is *systematic error* (i.e., bias) in saccade targeting, such that the amplitude of all intended saccades is shifted closer to a *preferred saccade length*, which is seven characters in E-Z Reader and about five characters in SWIFT.[3] In both models, the amount of bias is proportional to the difference between the intended and preferred saccade lengths, and thus there is more bias for saccades intended to be especially long or short. For example, in E-Z Reader, a saccade intended to be nine characters might be pulled towards the preferred length of seven characters by a character, producing a saccade effectively targeted for a position eight characters away, yet a saccade intended for a position 11 characters would have twice the bias and end up targeting a position nine characters away.

The concept of systematic error is not descended from the notion of efficiently getting disambiguating visual input; nevertheless, it functions in the models to produce some of the same effects. In combination with the functional target of saccades in these models always being the centre of the word, systematic error allows the model to reproduce a number of aspects of human reading behaviour that we earlier argued could be explained as readers moving their eyes to efficiently obtain disambiguating visual input. One aspect of human behaviour that systematic error helps these models to capture is the shift in the peak of the initial landing site distributions caused by varying the saccade's launch site. Specifically, the closer the previous

[3] Technically, in SWIFT, the preferred saccade length is different for progressive nonrefixations, progressive refixations, regressive refixations, and regressive nonrefixations.

fixation is to the word, the further right the peak shifts, an effect we suggested could be understood as allowing for more efficient word identification because fixations closer to a word provide more parafoveal preview of the word's initial letters. Systematic error allows E-Z Reader and SWIFT to reproduce these effects because the amount of undershoot (or overshoot) grows as the distance to the target word's centre becomes larger (smaller). For example, if the centre of the targeted word is seven characters from the current point of fixation, then the effective saccade target will also be seven characters; but if the current fixation is further away from the centre of the targeted word, perhaps 11 characters, a position two characters left of the word's centre will be targeted. Another aspect of human behaviour that E-Z Reader and SWIFT can reproduce by depending on systematic error concerns refixations. For human readers, refixations launched from the beginning of a word typically take the eyes to the word's ending (Rayner et al., 1996), an effect that we suggested could be understood as allowing for more efficient word identification because, after a fixation at the beginning, most of the visual information about the word that has not yet been obtained is at the end. In E-Z Reader and SWIFT, a refixation initiated from the beginning of a word, like all saccades, will be targeted to the centre of the word and subject to systematic error. Except for very long words, the distance of the current fixation from the centre of the word (the intended saccade target) will be less than the preferred saccade distance, and thus, systematic error will cause the eyes to overshoot the word's centre, resulting in a fixation on the word's end. In each of these cases, systematic error allows the models to reproduce aspects of human reading behaviour that could otherwise be naturally explained on an efficient visual identification account, despite the fact that under the models' simplified assumptions about word identification, these behaviours actually result in less—not more—efficient identification.

Clearly, this approach of including within a model of reading only simplified assumptions about the word identification process has been very productive, and can reproduce a number of human reading phenomena. The goals of using such a simplified model of word recognition, according to Reichle, Rayner, and Pollatsek (2003, section R3) were to make the models (1) more transparent and (2) simple enough computationally to evaluate on large corpora. With respect to these criteria, this program has been a success. It is important, however, to remember the limitations of such an approach. Perhaps the most obvious limitation of using a simplified word identification model that does not include any notion of disambiguating visual input is that the model treats all words similarly, and cannot distinguish between words with different properties. For example, words that have disambiguating information in different places, as described earlier, will not be predicted to differ, and for the same reason,

such models cannot reproduce orthographic neighbourhood effects (Pollatsek, Perea, & Binder, 1999). Certainly, this will decrease the model's ability to make accurate predictions on cases in which these factors are relevant. Perhaps more dangerous, however, is the implicit assumption that there is no variance between words with regard to these properties, which may also harm the model's performance in the aggregate. For example, reading a series of words whose optimal viewing positions vary wildly but are on average in the centre of the word should yield very different reading behaviour than reading a series of words whose optimal viewing positions are each exactly at the centre. Thus, one must be somewhat cautious when interpreting the predictions of a model of eye movements in reading that does not incorporate a model of word recognition from disambiguating visual input. One motivation for pursuing a model of eye movement control in reading that does incorporate such a model of word recognition, then, is to ensure that the simplified model of word recognition is not distorting the model's predictions too badly.

Perhaps the more exciting reason to pursue models of eye movement control in reading that include models of word recognition from disambiguating visual input is to test and sharpen our intuitions for what would constitute efficient reading behaviour for word identification. We have already presented intuitions as to how a number of eye movement phenomena might be explained in terms of efficient reading behaviour for word identification. For example, when we suggested that we might understand the fact that word skipping rates decrease as word length increases, we made a number of conjectures: (1) The amount of parafoveal preview available about a word is relatively independent of its length and (2) the probability of being able to identify a word from a fixed amount of parafoveal preview decreases as word length increases. Although these both seem to be reasonable hypotheses, either or both of them may in fact be incorrect. Thus, if we are to claim that this phenomenon can be explained as resulting from efficient reading behaviour for word identification, we must first verify that this behaviour actually does result from efficient reading for this goal. One way to make such a demonstration is to perform simulations with an implemented model of eye movements in reading that performs word identification from visual input and moves its eyes efficiently given this goal. If the behaviour of such a model reproduces the eye movement phenomenon in question, then this constitutes evidence that the phenomenon can be understood as arising from efficient reading for word identification from visual input. Thus, if we are to argue that many eye movement phenomena in reading should be understood as naturally resulting from an efficient visual identification account, one crucial piece of the argument is a demonstration that a model efficiently identifying words from visual input can actually reproduce those phenomena.

MR. CHIPS

For many eye movement phenomena, the intuition that they can be explained as resulting from an efficient visual identification account has already been verified by the only extant model of eye movement control in reading to incorporate a model of word identification from visual input, Mr. Chips (Legge et al., 1997, 2002).[4] This model makes the simplifying assumption that all fixations are of approximately equal length, and thus give equal quality visual input, which consists of the (veridical) identities of the nine characters around the point of fixation as well as peripheral information about word boundaries in the four character positions on each side of this range. Mr. Chips uses this visual input to identify words in series, one at a time, by continuing to obtain visual input about a word until it has eliminated all possible identities of the word except one. The model uses this formalization of word identification to plan saccades according to the following heuristic. It first calculates a probability distribution over possible identities of the word currently being identified by combining the visual input obtained thus far with its knowledge of how likely each word is in the language. (Note that this means that the model does not make use of the prior linguistic context to identify words.) Then, the model targets its next saccade to the position that it calculates will provide the most additional information about the identity of the word (formally, the position expected to minimize the entropy in its distribution over possible word identities), taking into account its saccade motor error.[5] The model continues to make saccades in this manner until the current word is identified with complete certainty, and then it focuses on identifying the next word (or the one after that if the next word can already be uniquely identified based on previous visual input).

Despite its overly simplistic model of visual input, this model can reproduce a number of facets of human eye movement behaviour (Legge et al., 2002), many of which verify intuitions presented earlier that certain

[4] The model of eye movement control in reading other than Mr. Chips that comes closest to this goal is Glenmore (Reilly & Radach, 2006), which incorporates a connectionist model of letter and word activation that bears some similarity to interactive activation models of word identification (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Crucially for our purposes, however, Glenmore differs from interactive activation models of word recognition in only having a single letter node for each character position (rather than multiple possible letter identities) and a single word node for each word (rather than multiple possible word identities). That is, the model entertains no other candidate letter or word identities other than the correct one, and thus cannot perform word identification.

[5] See Appendix A for formal details of how this algorithm works.

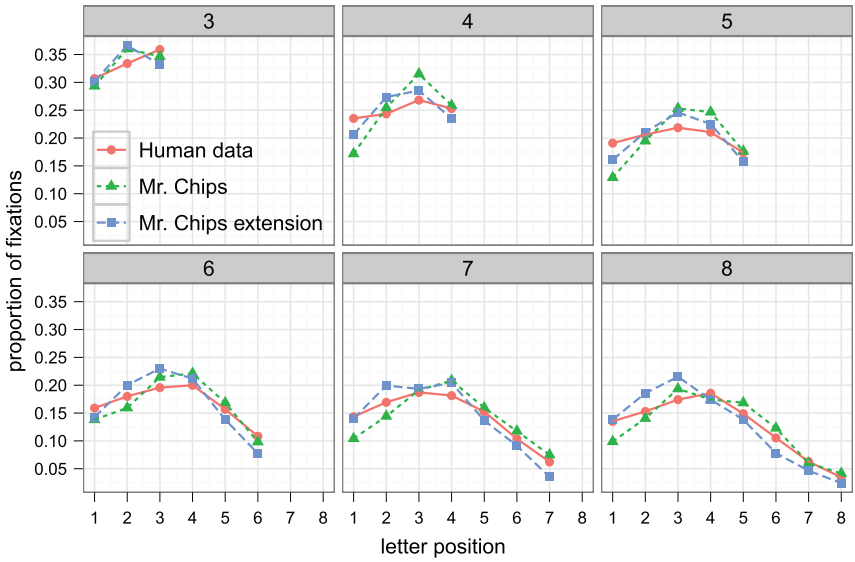

Figure 1. Proportion of first fixations as a function of letter position within words of lengths 3–8 in the behaviour of humans, the Mr. Chips model, and our extended version of the Mr. Chips model. We extracted the human data from the Dundee Corpus of eye movements (Kennedy & Pynte, 2005), and measured the Mr. Chips model and our extension of the model using our own implementation. The extended version shown was parameterized to use context and a 90% identification criterion. To view this figure in colour, please see the online issue of the Journal.

phenomena can be understood as resulting from an efficient visual identification account.[6] As one important example, the model can reproduce the preferred landing position effect, in which the most likely position within words to be initially fixated is at or just left of centre (for English). Figure 1 shows the distribution of initial landing positions on words that are between three and eight characters long for the Mr. Chips model and our extension of the Mr. Chips model (which we describe later), both measured using our own implementation, as well as human data for comparison (calculated from the Dundee Corpus of eye movements; Kennedy & Pynte, 2005). As can be seen from Figure 1, the distribution of initial fixations on words produced by both models (and humans) peaks at or just left of the centre of words (for words of at least four characters). Because the algorithm used by the Mr. Chips model to select saccade targets works by choosing the position

[6] Note, however, that none of the following simulation results we describe from the Mr. Chips model show sensitivity to the particular visual information within words. It seems reasonable given the model's algorithm for saccade target selection to suppose that the model will, e.g., be more likely to skip the endings of words whose beginnings uniquely identify the word, but simulations to verify this have not been performed.

that is expected to provide the most information about the word in question, this result corroborates O'Regan's (1981) idea that we can understand the preferred landing position effect as resulting from the fact that the position is on average the most useful one to identify words. Further, the result goes beyond that of Clark and O'Regan (1999), by demonstrating that this explanation works not just for single word recognition, but is still on average true of the reading of continuous text, despite the complications of parafoveal preview, linguistic context, and motor error.

As another example, recall the effect of launch site on the distribution of initial fixation positions on words, i.e., that the peak of the distribution shifts forwards (or backwards) as the saccade launch site becomes closer to (further away from) the word. We previously suggested that we might be able to understand this in terms of efficient visual identification: Given that positions closer to a word are likely to yield parafoveal information about the first characters of the word, the most useful visual information in this case that the reader has not yet obtained will be at the end of the word, yielding on average a peak in initial fixations that is further forward, closer to the end of the word. Legge et al. (2002) present results showing that the behaviour of the Mr. Chips model displays this effect, supporting the idea that the effect of launch site can be understood along these lines.

In addition to these effects on initial fixation positions, the model also verifies a number of other intuitions for how eye movement phenomena can be understood as resulting from an efficient visual identification account, including the effect of a word's length on the rate at which it is skipped and the effect of initial fixation position on refixation rate. The fact that a single model of eye movement control in reading that explicitly models the process of word identification and produces eye movements designed to maximize identification efficiency can reproduce such a range of phenomena provides substantial support for the notion that a wide range of reading phenomena can be understood as resulting from efficient visual identification.

THE EFFECT OF CONTEXT

One prominent eye movement phenomenon in continuous reading for which the Mr. Chips model does not provide an account, however, is the effect of linguistic context. It has been known at least since Morton (1964) that the basic effect of context is to allow for faster reading. Morton demonstrated that reading of contextualized text is 33% faster than reading random words. Under a framework in which readers are identifying the words in the text, the basic role of linguistic context is to provide a second source of information—in addition to the visual input—about the words' identities. For example, given only visual input that the first letter of a word is "c" one

is left quite uncertain about the identity of the word; but knowing that the preceding context is "The first thing I drink every morning is . . ." gives substantially more information about this word's identity.

Because context can give additional information for word identification, it may be suggested that a reader who uses this information for word identification can identify words (and thus read) more efficiently when contextual information is available than when it is not, explaining this effect also as resulting from efficient visual identification. In the Mr. Chips model, however, linguistic knowledge is taken to be simply word frequency information, and thus the model is unable to use context as an additional source of information for word identification, meaning that we cannot use the Mr. Chips model to evaluate whether (and how) this intuition would actually work out.

In the remainder of this paper, we describe two specific proposals for how context could increase reading speed under this framework, and then present an extension of the Mr. Chips model that can make use of context, which we use to evaluate these specific proposals. The results provide a case study in the utility of models that incorporate detailed models of word recognition: Showing that they can help to give insight into the ways in which eye movement phenomena in reading may (and might not) be understood as properties of efficient visual identification. Before we describe these proposals for how context effects may arise from efficient eye movements for word identification, we first describe effects of context in E-Z Reader and SWIFT.

EFFECTS OF CONTEXT IN MAJOR MODELS

One function of the linguistic context is to make a word in a text more predictable in some instances and less predictable in others. For example, "coffee" would be very predictable given the preceding context "The first thing I drink every morning is . . .", but it would be a more surprising continuation given a preceding context "Before bed, the last thing I do after brushing my teeth is drink a glass of . . .". These effects of predictability have been well studied in the reading literature, and it is known that when a word is more predictable given its preceding linguistic context, it is more likely to be skipped (Balota et al., 1985; Ehrlich & Rayner, 1981) and—when it is actually fixated—it is fixated for a shorter duration on average (Balota et al., 1985; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner & Well, 1996).

It is via such effects of predictability that E-Z Reader and SWIFT encode effects of context. Specifically, each of these models explicitly builds in effects of predictability on the "word processing rate" functions, which in turn lead to more skipping and shorter fixations on highly predictable words.

In E-Z Reader, the time required to process a word is broken down into two components, called L_1 and L_2, and the time required for each of these components is given by functions that yield shorter times for words that are more predictable. Intentional skipping in the model requires that both L_1 and L_2 for a word complete prior to the word being fixated, which will mean more skipping of highly predictable words. Similarly, when a word is fixated, a saccade to leave the word starts being programmed when L_1 finishes, which will thus happen earlier on average for highly predictable words, yielding shorter fixation times.

In SWIFT, the situation is slightly more complicated. There, word processing is split into two stages called *preprocessing* and *completion*. As in E-Z Reader, completion is defined to be shorter when words are highly predictable, but for preprocessing this is not the case. Preprocessing is actually set to be longer for highly predictable words when they are being processed parafoveally (i.e., not being fixated) and does not vary as a function of predictability when the word is being fixated. The reason for this is that words in SWIFT are more likely to be fixated when parafoveal preprocessing is closer to completion. Thus, making parafoveal preprocessing slower for highly predictable words ensures that these words will have higher rates of being skipped. When these words are fixated, however, word processing is faster for more highly predictable words, meaning that fixation durations will be shorter.

The previous discussion makes it clear that the basic effects of predictability—fewer and longer fixations on highly predictable words—will be reproduced by E-Z Reader and SWIFT. From this description, however, it is not necessarily clear that the overall effect of context, i.e., speeding up reading, will be apparent in the behaviour of these models. To see why it is indeed the case that context speeds reading in these models, we must describe the relationship between predictability and lexical processing time in these models in more detail. In each model, the predictability of a word is formally defined to be the probability of the word in context, generally estimated by a Cloze task, and the lexical processing times are taken to be a linear function of this probability. Because this relationship is taken to be linear (and not, e.g., logit; Agresti, 2002), the only substantial differences in lexical processing rates between words will be between those with a relatively high probability in context and those that are close to zero. (That is, there will be no real differences between words that both have relatively small values of predictability, say .01 and .0001, despite the fact that there are multiple orders of magnitude difference between them.) If we constructed versions of E-Z Reader and SWIFT that did not make use of context, and instead used values for predictability that were proportional to word frequency, the overall effect then would be that there were fewer words that had relatively high predictability, which would have the effect of slowing

reading. Hence, E-Z Reader and SWIFT do indeed predict the overall effect that context has of speeding reading.

TWO PROPOSALS FOR THE EFFECT OF CONTEXT

By building the effects of predictability directly into the word processing rate functions, E-Z Reader and SWIFT can reproduce the basic effects of context. But insight into the reasons why these effects might occur on an efficient visual identification account requires additional analysis. There are at least two suggestions in the literature for the answer to this question.

The first possibility for how context could allow for faster reading behaviour in a framework such as Mr. Chips is given by the authors of the Mr. Chips model. Legge and colleagues (2002) suggest that the efficiency of reading in a model such as Mr. Chips (i.e., the average saccade size) is largely a function of the model's average uncertainty about each word's identity prior to obtaining any visual input about it (formalized as *entropy*: See Cover & Thomas, 2006). To support this notion, they construct a set of artificial languages with vocabularies of different sizes by subsampling from an English lexicon, and note that the average uncertainty about a word in those languages prior to receiving visual input about it is larger for the languages with larger lexica. They then use the Mr. Chips model to simulate reading of each of those languages, and demonstrate that the model's average saccade size is smaller for languages with larger lexica, and thus for larger uncertainty. Legge et al. conjecture that, because context also reduces uncertainty about a word's identity, it will naturally lead to behaviour with larger saccades. That is, context will enable the model to select better saccade targets from which visual input is more efficiently gathered to fully disambiguate each word because of its better prior knowledge of what the word is likely to be.

A second possibility for how context could allow for faster reading is that it may enable a reader to require less visual input to reach a given level of confidence about predictable words. For this intuition, we must imagine that instead of requiring that a word is disambiguated completely, a reader is satisfied with being 90% confident about the word's identity (i.e., believing there is a 10% chance that it is not that identity). In this case, we can say that the word is "identified" when the reader's confidence in a particular identity of the word reaches this 90% threshold. Now make the simplifying assumption that prior to obtaining any visual information about a word, the reader has veridical knowledge of the preceding context. In that case, the probability of the true identity of that word under the reader's beliefs is given by the word's predictability in context, which we will denote by π. Thus, the reader's initial confidence about the true identity of the word is given by π,

and—assuming correct identification—the reader will continue gathering visual input about the word until their confidence in that identity reaches the 90% threshold. On average, then, less visual input will be required to reach this threshold for words that are highly predictable in their contexts, because confidence about the true identity of the word begins at a higher level. Note that this argument does not hold if the confidence criterion is always 100%, because in this case, the same amount of visual input will be needed regardless of predictability (i.e., enough to completely rule out all other possible words). This intuition for the effect of context—that more predictable words need less visual input to reach a given level of confidence—is closely related to rational accounts of frequency effects in isolated word recognition (Moscoso del Prado Martín, 2008; Norris, 2006, 2009). For reading of words in context, however, it is to our knowledge relatively unexplored.

Each of these suggestions for how the use of context might allow a reader to increase their efficiency appear reasonable, but it is important to note that neither has been tested in an implemented model of eye movement control in reading. Thus, it could be the case that one or both of these explanations cannot in practice significantly increase the rate of reading. In order to gain more insight, then, into the reasons why context may speed reading, we test these two intuitions by building an extension of the Mr. Chips model.

EXTENDING MR. CHIPS

In order to use the Mr. Chips framework to test these two intuitions for why context might allow a reader to be more efficient, we must first extend the model in two ways: (1) To allow for the use of contextual information in word identification and (2) to allow for a confidence criterion below 100%. We alter the model to use contextual information by replacing its knowledge of word frequency statistics with a word bigram language model (Jurafsky & Martin, 2009), encoding the transition probabilities between each pair of words in the language.[7] We allow for a lower confidence criterion by changing the model of word identification so that it shifts its focus on to the next word whenever the model's posterior probability of some identity of the current word exceeds a flexible threshold α. Finally, these two changes require that we update the model's saccade targeting algorithm so that it still targets the position

[7] Although we do not believe that word bigram language models are good approximations of how human readers make use of context (and do not encode the type of longer range context effects that reading researchers typically study), we use them here because of their computational simplicity. Such a model gives a lower bound on the amount of benefit that humans might obtain from context, and is sufficient for understanding the qualitative effect of using context in reading.

expected to give the most information about the word given these two changes to the model. Details of the modifications are reported in Appendix B.

SIMULATION

We use our extension of the Mr. Chips model to test the two intuitions for the effect of context mentioned earlier. Recall that the first intuition suggests that using context will allow the model to more efficiently gather visual input for 100% identification. We can test this using the extended Mr. Chips model by fixing the confidence criterion α at 100%, and then comparing the model's reading efficiency when it uses context to when it has access only to frequency information (like the original version of Mr. Chips). The second intuition suggests that context functions by allowing the reader to read efficiently with less visual input about predictable words, an intuition that only holds if the confidence criterion is less than 100%. We can test this intuition by setting the confidence criterion below 100%, and then comparing the model's performance when using context to when using just frequency information. In this simulation, we test these two hypotheses and systematically explore the relationship between the effect of context and the confidence criterion by performing simulations with a range of models that vary in (1) whether they make use of context or only frequency information and (2) their confidence criterion. We evaluate the reading efficiency of the models with two measures: Average saccade size and proportion of words skipped.

Methods

We represent the language knowledge of the frequency-only model with a unigram language model and of the model with context with a word bigram language model. Both models were smoothed with Kneser-Ney under default parameters (Chen & Goodman, 1998; equivalent to add-δ smoothing for the unigram model). As in Legge et al. (2002), the models were trained on a 280,000 word corpus of *Grimm Brothers' Fairy Tales*, containing 7503 unique words. This corpus was normalized by Legge et al. to convert all letters to lowercase, remove all punctuation other than apostrophes, convert all numbers to their alphabetic equivalents, and remove all nonsense words.

We test models with context and models with only frequency information at a range of six levels of the confidence criterion: 90%, 95%, 99%, 99.9%, 99.99%, and 100%, a total of 12 models.[8] We evaluate each of these models

[8] The models at the extremes of this range—the 100% model without context and the 90% model with context—were used to demonstrate the landing position results in Figure 1.

by simulating the reading of two different 40,000 word texts. Following the procedure used by Legge et al. (2002), one text is artificially generated by the model's internal language model, creating a situation under which the model has exact knowledge of the statistical regularities underlying the text. In addition, in order to ensure that no artificial properties of these texts influence the results, we evaluate each model on naturalistic text—the first 40,000 words of *Grimm Brothers' Fairy Tales* (i.e., the text on which the models' language knowledge is based.)

Results

Figure 2 reports the average saccade sizes of each of the twelve models on each of the two types of text and Figure 3 reports the proportion of words skipped.[9] Perhaps the most striking pattern from the four graphs is that they all look very similar. Comparing models with and without context when using a 100% confidence criterion (the furthest right points in each graph) reveals a striking disconfirmation of the predictions made by the intuition that adding context to the Mr. Chips model will substantially increase its efficiency at identifying words with complete certainty. In fact, the simulation results show that the model using context is slightly less efficient: It reads slightly more slowly when reading natural text, and skips slightly fewer words on both types of text. Thus, at least in this framework, it does not appear that using context can increase efficiency at identifying words with complete certainty.

The situation is quite different, however, when we examine the models that use confidence criteria below 100%. Here there is an interesting interactive pattern, in which the use of context gives more benefit to models with a lower criterion. This pattern of results verifies the intuition that one way in which context can facilitate more efficient reading is through allowing words to reach a confidence criterion with less visual

[9] Although it is orthogonal to our main point here, in which we only use average saccade size and word skipping rate as indices of reading speed, one may ask to what extent the model's saccade size and skip rate resemble human reading behaviour. Unfortunately, both of these measures vary as a function of a number of variables (e.g., text difficulty), so it is difficult to draw a precise comparison. That said, the values produced by the model do appear to be within the usual human range. Rayner (1998) gives the mean saccade size when reading English to be 7–9 characters, a range some of the models we tested fall into (only those with the lower confidence criteria, and more models with context than without). Regarding human readers' overall word skipping rates in English, a sample of empirical estimates is given by Rayner and McConkie (1976), who report a rate of 51%, Vitu, O'Regan, Inhoff, and Topolski (1995), who report 42%, McDonald and Shillcock (2003), who report 44%, and Greenberg, Inhoff, and Weger (2006), who report 40%, a comparable range to that of our models (39–46%).

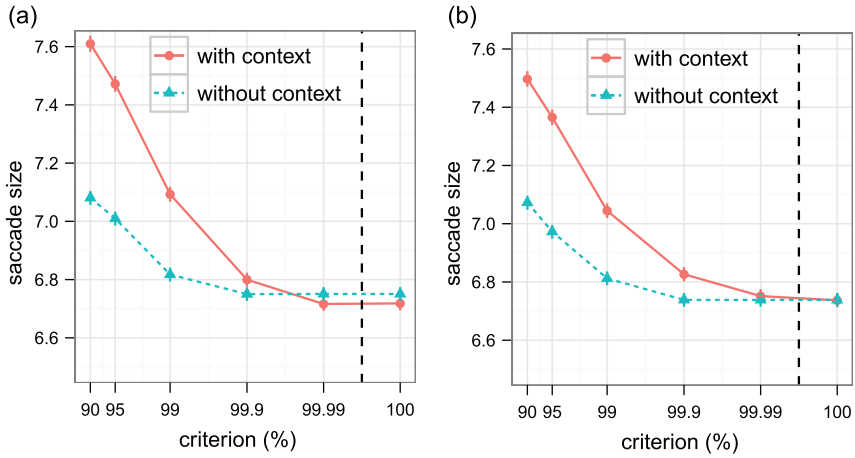

Figure 2. Effects of context and confidence criterion on mean saccade size in the extended Mr. Chips model, evaluated when reading natural text (a) and artificial text generated according to the model's knowledge of language (b). Confidence criteria are plotted on a logit-transformed $\left(\log\frac{p}{1-p}\right)$ scale, except the criterion of 100%, which is plotted as the rightmost value. Mean saccade sizes are plotted with 95% confidence intervals. To view this figure in colour, please see the online issue of the Journal.

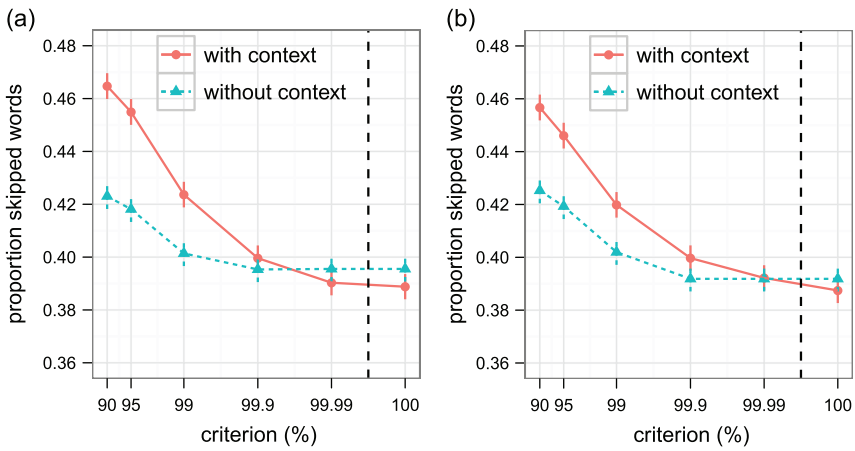

Figure 3. Effects of context and confidence criterion on the proportion of skipped words in the extended Mr. Chips model, evaluated when reading natural text (a) and artificial text generated according to the model's knowledge of language (b). Confidence criteria are plotted on a logit-transformed $\left(\log\frac{p}{1-p}\right)$ scale, except the criterion of 100%, which is plotted as the rightmost value. The proportions of skipped words are plotted with 95% binomial confidence intervals. To view this figure in colour, please see the online issue of the Journal.

input. Furthermore, it appears that the lower the criterion the more help context can give.

Discussion

In summary, the pattern of results presented here provides support for the notion that context can make reading more efficient for a reader who is content with substantially less than 100% certainty about the identity of each word, by allowing the reader to become confident about the identities of words with, on average, less visual input. It provides no evidence, however, for the notion that context can increase reading efficiency by allowing the reader to more efficiently get visual input for full word disambiguation. Of course, it could be the case that this result depends on the details of the Mr. Chips framework or of its algorithm for selecting saccade targets, and that contextual information may be useful to a different model in helping the reader to more efficiently gather visual input for full word disambiguation. However, these results with the Mr. Chips model provide clear evidence for one intuition of how context can increase reading efficiency and provide no support for the other. This is an example of one way in which a model of eye movement control in reading that incorporates a model of word identification from visual input can be especially useful: In showing that, of two reasonable intuitions for how context might affect reading behaviour, only one may help increase reading efficiency in practice.

GENERAL DISCUSSION

In this paper, we presented an argument for the incorporation of models of word identification from visual input into models of eye movement control in reading. We first described the extensive evidence that the process of isolated word recognition depends crucially on the particular visual input that the eyes receive about a particular word, as well as more limited evidence that this is also the case for the identification of words embedded in text. We also gave intuitions for how a number of eye movement phenomena from reading can be understood as naturally arising as part of a reader efficiently gathering disambiguating visual input for word identification, which we termed an efficient visual identification account. We noted that major models do not incorporate detailed models of word identification, but rather make simple assumptions about how the process of word identification works. Although these assumptions are enough to reproduce

some of these phenomena, they are not enough to reproduce them all. More crucially for the prospect of understanding eye movement behaviour as efficient visual identification, these models cannot be used to gain insight into the reasons why efficiently gathering visual input might give rise to many eye movement phenomena we observe.

We described the Mr. Chips model, the only extant model of eye movement control in reading that does incorporate a model of word identification from visual input, and noted that it verifies many of the intuitions described here by reproducing many aspects of human eye movement behaviour from a principle of moving the eyes to the most informative particular place to identify a particular word. We then provided an example of how a model of eye movement control in reading that incorporates a model of word identification from visual input could be used to test new intuitions for how the use of linguistic context could affect reading on this account. Accomplishing this required the creation of an extended version of the Mr. Chips model that could (1) make use of linguistic context and (2) be content with less than 100% certainty about the identities of words. Simulations with this extended model revealed that context can speed reading only in the case that readers do not require 100% certainty about the identity of each word, providing support for one intuition about how context could affect reading behaviour, but no evidence for the other.

We have argued that there is good evidence that many aspects of reading behaviour can be well understood in an efficient visual identification account, in which readers move their eyes to efficiently obtain visual input for word disambiguation. However, for this account to be viable, it must be demonstrated that it can provide an explanation for many more aspects of reading behaviour. The only way to provide such demonstrations is through the use of models of eye movement control in reading that incorporate a model of word identification from visual input, such as Mr. Chips.

Our efficient visual identification account is closely related to rational models of cognition (Anderson, 1990) and to Marr's (1982) computational level of analysis. These paradigms assume that the cognitive system optimizes the behaviour of an agent given the agent's goals and the constraints posed by the task, and seek to understand human behaviour in terms of the efficient ways for agents to perform that task, given the constraints. In the case of our extension of Mr. Chips, the model's goal is defined to be efficient serial identification of each word to a given level of confidence and the task constraints are given by the model's language knowledge, its visual input system, and its motor error. Mr. Chips' algorithm

for saccade target selection represents an efficient solution to this problem.[10] In rational analysis, the efficient solution to the task is compared to human behaviour, and if its predictions are found to be incorrect, it is taken to suggest that either the agent's assumed goal or the task constraints should be revised.

Of course, we are not the first to argue for the benefits of incorporating realistic models of word identification into models of eye movement control in reading (e.g., Grainger, 2003; Huestegge, Grainger, & Radach, 2003). In fact, in this Special Issue, Reichle, Rayner, and Pollatsek (this issue 2012) reach a very compatible conclusion. They describe a sort of rational analysis performed using their E-Z Reader model, in which they run simulations with the model to determine the most efficient time for a reader to initiate a saccade to leave a word. Formally, word identification in E-Z Reader is broken up into two serial stages, first L_1 and then L_2, and it is the completion of L_1 that triggers initiation of a saccade to leave a word. Given that there is a delay between the initiation of a saccade and its execution, initiating a saccade when L_1 completes will often mean executing the saccade around the time L_2 completes, and that the word is fully processed before the eyes leave it. Reichle et al. performed simulations for a range of values of the total word processing time ($L_1 + L_2$), and tested the effect on reading speed of varying the proportion of this total comprised by L_1, i.e., the proportion of total word processing time that has transpired when a saccade to leave the word is initiated. The results of the simulations reveal that with the current version of E-Z Reader, the most efficient time for a reader to initiate a saccade to leave a word is immediately; that is, reading speed increases monotonically as the saccade to leave a word is initiated earlier. As in rational analysis, Reichle et al. suggest that the fact that this conclusion does not comport with human reading behaviour (as human readers do not immediately initiate saccades to leave words upon landing) suggests that the constraints on reading imposed by their model are misspecified. The two main possibilities they highlight for how their model constraints may be misspecified both relate to the assumptions the model makes about the word identification process, and specifically about the role of visual information within it. The first concerns the fact that in E-Z Reader, the speed of L_1 is sensitive to visual information (decreasing with the average distance of each letter in the word from the fovea) but the speed of L_2 is unaffected by visual properties. Given this, one interpretation

[10] It should be noted, however, that Mr. Chips' algorithm is not necessarily the most efficient solution to the problem. For example, as each saccade is planned to maximize the information obtained about the current word, ignoring any information that might be obtained about the next word, this algorithm can be somewhat short-sighted as a way of minimizing the time required to identify the entire text.

of their findings is that while initiating a saccade to leave a word too early means that the reader will still be processing the word after its eyes have left, there is no penalty for this in the model, as L_2 is insensitive to visual information. Reichle et al. suggest that a natural solution to this problem in E-Z Reader would be to make L_2 sensitive to visual information (perhaps in the same way as L_1), meaning that the entire word identification process in E-Z Reader would be sensitive to visual information, bringing the model more in line with an efficient visual identification account of reading. The second possible alternative Reichle et al. suggest for how the constraints in E-Z Reader may need to be revised concerns the notion of word misidentification, which recent work has indicated may play an important role in eye movement behaviour (Levy, Bicknell, Slattery, & Rayner, 2009; Slattery, 2009). Reichle et al. speculate that if the part of the word identification process that is sensitive to visual information (L_1) is performed too rapidly, it could lead to an increase in the number of misidentified words, which would have their own penalty on reading efficiency by causing integration failure downstream. They note, however, that modelling word misidentification is ''outside of the E-Z Reader model's theoretical scope'' (p. XXX). We argue that both of these possibilities for what E-Z Reader is missing given by Reichle et al.—a larger role for visual input throughout the word identification process and a nonnegligible probability of word misidentification dependent on visual input—point to the lack of a model of word identification from visual input in the model. Indeed, these possibilities are natural consequences of an efficient identification account.

As a model of the efficient identification account, the Mr. Chips framework has a limitation, however, that prevents it from being able to reproduce key aspects of human reading behaviour. Namely, it cannot make predictions for the durations of fixations, but rather only their locations. In many domains, it is common for models of eye movement control to only model the *where* or *when* components of eye movements. For example, most models of eye movements in visual search (e.g., Najemnik & Geisler, 2005) and scene viewing (e.g., Itti & Koch, 2000) only model fixation locations, while Nuthmann, Smith, Engbert, and Henderson (2010) present a scene viewing model only of fixation durations (see also Nuthmann & Henderson, this issue 2012). Mr. Chips is unique, however, among models of eye movement control in reading in not modelling fixation durations, and knowledge about effects on the durations of fixations comprise a large amount (if not the majority) of our knowledge of eye movements in reading. The ultimate reason why Mr. Chips is unable to model fixation durations derives from the nature of visual input in the model. After a single timestep fixating a particular location, the model receives veridical information about the identities of the nine characters surrounding the point of fixation.

Because it would obtain no additional visual information by spending another timestep fixating that location, there is no reason for the model ever to do so, resulting in all its fixations being of equal duration (i.e., one timestep). In order to make predictions for variable fixation durations, then, there must be some reason why it would be efficient to spend more than one timestep fixating a particular location. One proposed remedy for this problem is to make the visual input stochastic, i.e., not veridical letter identities, so that the model can choose to fixate a position longer to obtain higher quality visual input, an approach explored by Bicknell and Levy (2010). Another possibility would be to leave visual input veridical, but only give the reader a particular number of letter identities on each time step. For example, there could be an expanding visual window around the point of fixation. Whichever of these methods is used, allowing the model to make predictions for durations will also necessitate adding other complications to the model, such as the time it takes to plan and execute a saccade. This appears, however, to be a necessary step toward allowing for the understanding of a much wider range of eye movement phenomena as resulting from efficient visual identification.

Another advantage of modelling eye movements in reading as a process of obtaining the most useful visual information for the task is that there are analogous models of eye movements in a number of other domains. For example, Najemnik and Geisler (2005, 2008) show that eye movements in a visual search task are well modelled as being targeted to obtain the most useful disambiguating visual input about which location in an array contains the target (see also Butko & Movellan, 2010; Zelinsky, 2008, this issue 2012). Similarly, Itti and Baldi (2009) report that humans preferentially move their eyes to locations that are especially informative in scene perception (see also Kanan, Tong, Zhang, & Cottrell, 2009; Torralba, Oliva, Castelhano, & Henderson, 2006; Zhang, Tong, Marks, Shan, & Cottrell, 2008) and Renninger, Verghese, and Coughlan (2007) show that eye movements in a shape discrimination task are well described as maximizing the total information gained about the shape. Even beyond the wide applicability of this framing in terms of efficiently obtaining information, the parallels in practice between domains when framed this way are striking. For example, whereas in this paper we highlighted the importance for identifying a word of the linguistic context around it, in another paper in this Special Issue, Marat and Itti (this issue 2012) demonstrate the importance for identifying an object in a scene of using the context around it. Investigating the extent to which we can understand eye movements in reading as efficiently gathering visual input for word identification, then, allows for a substantial amount of interaction with a range of eye movement tasks.

Of course, striving to explain eye movement behaviour across a range of tasks using a single framework is not a goal unique to the efficient visual

identification account. Three of the other papers in this Special Issue (each from a different perspective) also seek to model eye movement behaviour across tasks using a single model (Nuthmann & Henderson, this issue 2012; Reichle et al., this issue 2012; Schad & Engbert, this issue 2012). The strategy in doing so is similar across all three: They each fit the parameters of their model to the data from each task separately, and show that the resulting models (with task specific parameters) reproduce a number of the patterns in the empirical data. These results are important in demonstrating that each of the model frameworks can be used to understand eye movement behaviour in more than one task, but they also raise the question of how to interpret the differences in model parameters across tasks. That is, while these simulations reveal some underlying similarity between tasks, it is still unclear why eye movements look one way in one task and another way in another task. One advantage of rational approaches such as the efficient visual identification account is that they do not encounter this problem. Because eye movement behaviour is understood as being performed to efficiently achieve the agent's goal in the task, given relevant task constraints, it should change in systematic ways as the task goal and constraints change. For example, modifying a model like Mr. Chips to perform visual search for a particular word rather than word identification would involve changing the model's goal: Instead of trying to identify each word, the model would try to determine whether or not each word was the target word. This slightly modified goal would lead to a slightly modified saccade targeting algorithm, in which instead of moving the eyes to the position that will provide the most information about the identity of the current word, the eyes would be targeted to the position that will provide the most information about whether or not the word is the target word. Similarly, the criterion for moving on to the next word would be a function of the model's confidence in whether or not the current word was the target. That is, modifying Mr. Chips for visual search would only require changing its goal and working out the implications of that new goal for the model's algorithm. As the other task constraints—such as the models of language knowledge, the visual input system, and motor error—would not need to be changed, the differences between the model's behaviour in the two tasks would be interpretable as arising from the differences in the nature of efficient performance in the two tasks.

The efficient visual identification account we have proposed (and rational accounts of eye movement behaviour more generally) implicitly assumes that the agent can directly control when and where to move the eyes. Two papers in this Special Issue, however, provide evidence questioning each of these assumptions (Nuthmann & Henderson, this issue 2012; Zelinsky, this issue 2012). Zelinsky provides evidence from visual search tasks showing that a number of fixations land at positions in

between possible target locations, and he interprets this result as resulting from an intrinsic constraint of the motor system. Specifically, he presents a model (TAM; Zelinsky, 2008) in which a number of possible motor command vectors are entertained, and during saccade planning, this set of vectors is iteratively pruned, removing at each step the least desirable motor command. When this process runs to completion, the saccade will be targeted to the most desirable location. In many cases, however, this process will not finish prior to the saccade's launch. When this occurs, the saccade is sent to the spatial average of the remaining motor command vectors, yielding fixations targeted to locations exactly in between objects of interest. One possible objection to this interpretation of the results is the possibility that locations in between objects of interest may sometimes be the most desirable place to move the eyes. For example, Najemnik and Geisler's (2005, 2008) rational model of visual search also produces fixations between objects of interest, which are designed to gather some information about both of them. Zelinsky (this issue 2012) is aware of this objection and argues that the objects of interest in his experiments were spaced too far apart for a fixation halfway in between the two to provide useful information about either object. If this interpretation is correct, his results provide evidence that in some cases agents cannot directly control where their saccades are targeted. Similarly, Nuthmann and Henderson (this issue 2012) provides evidence from both reading and scene viewing tasks suggesting that agents cannot always directly control the duration of their fixations. Using an experimental paradigm in which at the beginning of one-sixth of fixations, the stimulus disappears, and only reappears again after a random amount of time, Nuthmann and Henderson show that a number of saccades are launched prior to stimulus onset, while other saccades appear to wait until the stimulus appears and is processed. They interpret this result as providing evidence for a model such as CRISP (Nuthmann et al., 2010), in which there is an autonomous timer that launches saccades. This timer is loosely coupled to cognition, but it is not under direct cognitive control. The interpretation of their experimental results, then, is that on some proportion of the trials on which the stimulus is delayed, a saccade is launched by the autonomous timer prior to stimulus onset despite cognitive processing from that location not having yet begun. At first sight, each of these results may be taken to provide evidence against a rational approach in which cognition moves the eyes to perform a given task most efficiently, as indeed, cognition cannot do so if eye movements are not under its control. However, recall that in the rational approach, each task has a particular set of task constraints, so each of these conclusions could be incorporated into a rational model of eye movements as motor constraints on the task, and the nature of efficient solutions to the task would change accordingly. Determining precisely what the motor

constraints are on eye movement control, then, represents a key part of determining efficient eye movement behaviour, and is an important direction for future research. Given the relevant task constraints, however, rational models of eye movement control promise to lead not only to new predictions and insights for the understanding of eye movements in reading, but to allow for a unified understanding of eye movement behaviour across all domains.

REFERENCES

Agresti, A. (2002). *Categorical data analysis* (2nd ed). New York, NY: Wiley.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364–390.

Bicknell, K., & Levy, R. (2010). A rational model of eye movement control in reading. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics (ACL)* (pp. 1168–1178). Uppsala, Sweden: Association for Computational Linguistics.

Butko, N. J., & Movellan, J. R. (2010). Infomax control of eye movements. *IEEE Transactions on Autonomous Mental Development*, *2*, 91–107.

Chen, S. F., & Goodman, J. (1998). *An empirical study of smoothing techniques for language modeling* (Tech. Rep. No. TR-10-98). Cambridge, MA: Computer Science Group, Harvard University.

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

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed). Hoboken, NJ: Wiley.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*, 621–636.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813.

Grainger, J. (2003). Moving eyes and reading words: How can a computational model combine the two? In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 457–470). Amsterdam, The Netherlands: Elsevier.

Greenberg, S. N., Inhoff, A. W., & Weger, U. W. (2006). The impact of letter detection on eye movement patterns during reading: Reconsidering lexical analysis in connected text as a function of task. *Quarterly Journal of Experimental Psychology*, *59*, 987–995.

Holmes, V. M., & O'Regan, J. K. (1987). Decomposing French words. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 459–466). Amsterdam, The Netherlands: North-Holland.

Huestegge, L., Grainger, J., & Radach, R. (2003). Visual word recognition and oculomotor control in reading. *Behavioral and Brain Sciences*, *26*, 487–488.

Hyönä, 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.

Hyönä, J., Niemi, P., & Underwood, G. (1989). Reading long words embedded in sentences: Informativeness of word halves affects eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 142–152.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *29*, 1295–1306.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed). Upper Saddle River, NJ: Prentice Hall.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*, 979–1003.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*, 153–168.

Legge, G. E., Hooven, T. A., Klitz, T. S., Mansfield, J. S., & Tjan, B. S. (2002). Mr. Chips 2002: New insights from an ideal-observer model of reading. *Vision Research*, *42*, 2219–2234.

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

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the USA, 106,* 21086–21090. Correction: *Proceedings of the National Academy of Sciences of the USA*, *107*, 5260.

Marat, S., & Itti, L. (2012). Influence of the amount of context learned for improving object classification when simultaneously learning object and contextual cues. *Visual Cognition*, *20*, 581–603.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W.H. Freeman.

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

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.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, *43*, 1735–1751.

Morris, R. K., Rayner, K., & Pollatsek, A. (1990). Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 268–281.

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.

Moscoso del Prado Martín, F. (2008). A fully analytic model of the visual lexical decision task. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 1035–1040). Austin, TX: Cognitive Science Society.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Science*, *434*, 387–391.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, *8*, 1–14.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*, 327–357.

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review*, *116*, 207–219.

Nuthmann, A., & Henderson, J. M. (2012). Using CRISP to model global characteristics of fixation durations in scene viewing and reading with a common mechanism. *Visual Cognition*, *20*, 457–494.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*, 382–405.

O'Regan, J. K. (1981). The "convenient viewing position" hypothesis. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 289–298). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

O'Regan, J. K., & Lévy-Schoen, A. (1987). Eye-movement strategy and tactics in word recognition and reading. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 363–383). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillère, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 250–257.

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

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

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1–56.

Pynte, J., Kennedy, A., & Murray, W. S. (1991). Within-word inspection strategies in continuous reading: Time course of perceptual, lexical, and contextual processes. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 458–470.

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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K. (2009). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 720–732.

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., 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.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin and Review*, *3*, 504–509.

Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of fixation in reading. *Perception and Psychophysics*, *31*, 537–550.

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.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445–526.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2012). Eye movements in reading versus non-reading tasks: Using E-Z Reader to understand the role of word/stimulus familiarity. *Visual Cognition*, *20*, 360–390.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin and Review*, *16*, 1–21.

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*, 34–55.

Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, *7*, 1–17.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60–94.

Schad, D., & Engbert, R. (2012). The zoom lens of attention: Simulating shuffled versus normal text reading using the SWIFT model. *Visual Cognition*, *20*, 391–421.

Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1969–1975.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*, 766–786.

Underwood, G., Bloomfield, R., & Clews, S. (1988). Information influences the pattern of eye fixations during sentence comprehension. *Perception*, *17*, 267–278.

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.

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception and Psychophysics*, *57*, 352–364.

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception and Psychophysics*, *47*, 583–600.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, *115*, 787–835.

Zelinsky, G. J. (2012). TAM: Explaining off-object fixations and central fixation tendencies as effects of population averaging during search. *Visual Cognition*, *20*, 515–545.

Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision*, *8*, 1–20.

Manuscript received August 2011
Manuscript accepted February 2012
First published online May 2012

APPENDIX A: DETAILS OF THE MR. CHIPS SACCADE TARGETING ALGORITHM

We now give the algorithm that the Mr. Chips model uses to select the intended target for the next saccade. First, note that given the visual input obtained by the model from the first to the ith fixation $\mathcal{I}_1^i$ and the word frequency information, the model can calculate the posterior probability of any possible identity of a word w that is consistent with the visual input by normalizing its probability from the language model by the total probability of all visually consistent identities,

$$p(w|\mathcal{I}_1^i) = \frac{\chi(\mathcal{I}_1^i, w)p(w)}{\sum_{w'} \chi(\mathcal{I}_1^i, w')p(w')} \tag{1}$$

where $\chi(\mathcal{I},w)$ is an indicator function with a value of 1 if w is consistent with the visual input $\mathcal{I}$ and 0 otherwise, and $p(w)$ is the probability of w under the language model.

To identify a given word, the model selects the saccade target $\hat{t}$ that, on average, will minimize the entropy in this distribution, i.e., that is expected to give the most information about the word's identity

$$\hat{t} = \arg\min_t E\left[H(w|\mathcal{I}_1^i)\big|t, \mathcal{I}_1^{i-1}\right] = \arg\min_t \sum_{\mathcal{I}_i} H(w|\mathcal{I}_1^i)p(\mathcal{I}_i|t, \mathcal{I}_1^{i-1}). \tag{2}$$

That is, the minimum can be found by calculating the entropy of the conditional distribution produced by each possible new input sequence and weighting those entropies by the probability of getting that input sequence given a choice of target location. In information theory (Cover & Thomas, 2006), the entropy of the conditional distribution $p(w|\mathcal{I}_1^i)$ is standardly defined as

$$H(w|\mathcal{I}_1^i) = -\sum_w p(w|\mathcal{I}_1^i)\log p(w|\mathcal{I}_1^i). \tag{3}$$

The second term in the formula for $\hat{t}$ is the probability of a particular visual input given a target location and previous input. Because of motor error in the execution of saccades, we must calculate this term by marginalizing over possible landing positions ℓ given a particular target position t

$$p(\mathcal{I}_i|t, \mathcal{I}_1^{i-1}) = \sum_\ell p(\ell|t)p(\mathcal{I}_i|\ell, \mathcal{I}_1^{i-1}) \tag{4}$$

where $p(\ell|t)$ is given by the motor error function. We then marginalize over possible words

$$p\left(\mathcal{I}_i|\ell,\mathcal{I}_1^{i-1}\right)=\sum_{w}\chi(\mathcal{I}_i,\ell,w)p\left(w|\mathcal{I}_1^{i-1}\right) \tag{5}$$

where $\chi(\mathcal{I},\ell,w)$ is an indicator function with a value of 1 if w is consistent with the visual input $\mathcal{I}$ obtained about w from position ℓ, and 0 otherwise. Putting these together, we have that $\hat{t}$ is selected as

$$\arg\min_{t}\sum_{\mathcal{I}_i}H\left(w|\mathcal{I}_1^{i}\right)\sum_{\ell}p(\ell|t)\sum_{w}p(\mathcal{I}_i|\ell,w)p\left(w|\mathcal{I}_1^{i-1}\right). \tag{6}$$

That is, we can calculate the expected entropy for each possible value of t by summing over all possible inputs, whose probabilities are given by summing over all possible identities of the word and landing positions.

APPENDIX B: DETAILS OF THE SACCADE TARGETING ALGORITHM IN OUR EXTENSION OF MR. CHIPS

Note that this section builds on, and thus presumes knowledge of, Appendix A. As in the original Mr. Chips model, at any given point in time, the model is working to identify one word. However, this revised model considers the goal of identifying this word achieved when the marginal probability of some identity for the word given the visual input exceeds a predefined threshold probability α. This flexibility requires the algorithm to be substantially modified to allow for uncertainty about previous word identities and the use of linguistic context. We denote the sequence of words as W, where the first word is W_1.

Because every word in Mr. Chips was identified with complete certainty, the model always knew precisely at which position the next word to be identified began, and its goal was always to identify this next word. Now that the model has uncertainty about the identities of previous words, however, the goal must be changed. In the revised model, the reader is always focused on some character position n, and its goal is to identify which word $W_{(n)}$ begins at that position (if any), with confidence exceeding α. Once the model has achieved this goal, it then chooses a new character position n via a procedure whose description we leave for later. To be explicit about this goal,

we slightly update our original equation for choosing $\hat{t}$, swapping w out for $W_{(n)}$

$$\hat{t} = \arg\min_{t} \sum_{\mathcal{I}_i} H\left(W_{(n)}|\mathcal{I}_1^i\right) p\left(\mathcal{I}_i|t, \mathcal{I}_1^{i-1}\right) \tag{7}$$

where the entropy is calculated assuming that some word does in fact begin at position n. The fact that our language model can now make use of linguistic context means that the equation for finding the probability of the current word given some visual input (Equation 1) must also be changed to marginalize over identities of the preceding words

$$p\left(W_{(n)}|\mathcal{I}_1^i\right) = \sum_{W_1^{(n)-1}} p\left(W_{(n)}\middle|\mathcal{I}_1^i, W_1^{(n)-1}\right) p\left(W_1^{(n)-1}|\mathcal{I}_1^i\right) \tag{8}$$

where $W_1{}^{(n)-1}$ denotes the range of words beginning with the first word of the sentence W_1 and extending through the word prior to the word beginning at position n. These probabilities of strings consistent with the visual input are again given probabilities according to their probability in the language model normalized by the probability of all other consistent strings (cf. Equation 1)

$$p\left(W|\mathcal{I}_1^i\right) = \frac{\chi\left(\mathcal{I}_1^i, W\right) p(W)}{\sum_{W} \chi\left(\mathcal{I}_1^i, W\right) p(W)}. \tag{9}$$

The second term in Equation 7 is expanded as in Mr. Chips by marginalizing over the possible landing position ℓ

$$p\left(\mathcal{I}_i|t, \mathcal{I}_1^{i-1}\right) = \sum_{\ell} p(\ell|t) p\left(\mathcal{I}_i|\ell, \mathcal{I}_1^{i-1}\right), \tag{10}$$

but now to incorporate information about the linguistic context, we must next marginalize over possible full sentence strings instead of possible words

$$p\left(\mathcal{I}_i|\ell, \mathcal{I}_1^{i-1}\right) = \sum_{W} \chi(\mathcal{I}_i, \ell, W) p\left(W|\mathcal{I}_1^{i-1}\right). \tag{11}$$

If we make the simplifying assumption that the model does not consider possible future input about words that are after $W_{(n)}$, this sum can again be

finitely computed for a given t by a relatively straightforward dynamic programming scheme. The range of possible values of t to search through also grows relative to Mr. Chips, because the model must consider not only any position that can give visual input about $W_{(n)}$ itself, but also positions that can give information about any position of uncertainty, since that may indirectly help to identify $W_{(n)}$ through linguistic context. In the case where the language model is an n-gram model, the probability of a word in context is a function only of the previous n-1 words. Thus, the minimum value of t that can contribute toward helping to identify $W_{(n)}$ cannot be further back than the most recent string of n-1 words for which the model has no residual uncertainty. Having established the method of selecting a saccade to identify $W_{(n)}$, we next give a description of the full algorithm of the model, including how to select n.

The model always begins reading by focusing on identifying $W_{(0)}$. Once the probability of some identity for $W_{(0)}$ is greater than α, all the possible identities of $W_{(0)}$ that have not been ruled out by visual input are combined into a set of possible "prefixes". Each of these prefixes has a conditional probability given the visual input, and each one predicts that the next word in the sentence begins at a particular position (i.e., two characters past the end of that string). Thus, the set of prefixes specify a probability distribution over the possible positions at which the next word begins. The model simply selects the most likely such position as the next character position n to focus on identifying $W_{(n)}$.

Now in the general case, the system has a set of prefixes together with their conditional probabilities given the visual input, and a position n, which it is trying to identity the word beginning at. It plans and executes saccades according to the formula for $\hat{t}$, and after getting each new piece of visual information, the model rules out not only possible candidates for the current word, but also possible prefix strings, and renormalizes both distributions. The model's attempt to identify $W_{(n)}$ can now end in one of two ways: (a) the model's confidence in some identity of $W_{(n)}$ exceeds the confidence threshold α or (b) the model eliminates all possible candidates for $W_{(n)}$ and thus knows that no word begins at that position. In the former case, the model creates all possible concatenations of prefixes ending two characters prior to $W_{(n)}$ (i.e., prefixes whose next word begins at n) with all the possible identities of $W_{(n)}$, and adds these new strings to the set of prefixes. Then, in both cases, it removes those original prefixes whose next word begins at n from the set. Note that this update of the list of prefixes leaves unaffected prefixes that are incompatible with a word beginning at position n, but still compatible with visual input. Finally, since the set of prefixes again gives a distribution over the starting position of the next word, the model selects the most likely new n and the cycle continues.