

# The Effect of Plausibility on Eye Movements in Reading

Keith Rayner

University of Massachusetts, Amherst

Tessa Warren

University of Pittsburgh

Barbara J. Juhasz

University of Massachusetts, Amherst

Simon P. Livversedge

University of Durham

Readers' eye movements were monitored as they read sentences describing events in which an individual performed an action with an implement. The noun phrase arguments of the verbs in the sentences were such that when thematic assignment occurred at the critical target word, the sentence was plausible (likely theme), implausible (unlikely theme), or anomalous (an inappropriate theme). Whereas the target word in the anomalous condition provided evidence of immediate disruption, the effect of the target word in the implausible condition was considerably delayed. The results thus indicate that when a word is anomalous, it has an immediate effect on eye movements, but that the effect of implausibility is not as immediate.

The amount of time that readers spend looking at a word is influenced by the ease or difficulty associated with processing the word (Livversedge & Findlay, 2000; Rayner, 1998). This has been shown by studies (see Rayner, 1998) demonstrating that readers spend more time looking at low-frequency words than they spend looking at high-frequency words (Inhoff & Rayner, 1986; Rayner & Duffy, 1986) and more time looking at words that are not constrained by the preceding context than they spend looking at highly predictable words (Ehrlich & Rayner, 1981; Rayner & Well, 1996). Models of eye-movement control in reading, like the E-Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003) and the saccade generation with inhibition by foveal targets (SWIFT) model (Engbert, Longtin, & Kliegl, 2002), use word frequency and predictability information as inputs to the system to predict fixation times and skipping rates (among other things). In this article, we examine the

role that plausibility has in influencing eye movements and fixation times on words.

There are two reasons why it is important to examine how plausibility influences eye movements in reading. First, as we noted above, E-Z Reader and SWIFT both rely heavily on word frequency and predictability as inputs for simulating fixation times in reading. It should be apparent that if there are other variables that influence fixation time on a word, to the extent that such variables can be identified, it should be possible to even better predict fixation times on words. For example, Juhasz and Rayner (2003, in press) recently demonstrated that a word's age of acquisition has an effect independent of frequency on fixation times. Clearly, the identification and systematic characterization of factors that independently influence fixation time on a word is an important objective in developing a comprehensive model of eye-movement control during reading.

The second reason why plausibility is an important variable is because plausibility manipulations have been widely used in the context of studies of sentence parsing. Here, the issue has typically been whether plausibility can override structural rules in initial parsing strategies. There is no question that plausibility has an effect on parsing; the issue is at what stage in the process (initial parsing decisions or later reanalysis) it has an effect. The data are indecisive on this issue, as some studies have reported that plausibility has little effect on the initial parsing process (Clifton, 1993; Clifton et al., 2003; Ferreira & Clifton, 1986; Pickering & Traxler, 1998; Rayner, Carlson, & Frazier, 1983), whereas others have found that it is used immediately to override structural preferences (Ni, Crain, & Shankweiler, 1996; Pickering & Traxler, 1998; Trueswell, Tanenhaus, & Garnsey, 1994). Thornton and MacDonald (2003) recently showed in a self-paced reading task that plausibility had an immediate effect on processing. However, self-paced reading is known to slow down the normal reading process (Rayner, 1998), so their results are not definitive with respect to the issue at hand. Our goal in the present research was

---

Keith Rayner and Barbara J. Juhasz, Department of Psychology, University of Massachusetts, Amherst; Tessa Warren, Department of Psychology, University of Pittsburgh; Simon P. Livversedge, Department of Psychology, University of Durham.

This research was supported by Grants HD17246 and HD26765 from the National Institutes of Health and by Grant S19168 from the Biotechnology and Biological Sciences Research Council. The work was undertaken when all of the authors were at the University of Massachusetts; Tessa Warren was supported by a Postdoctoral Traineeship on Grant MH016745 from the National Institute of Mental Health, Barbara J. Juhasz was supported on a Predoctoral Traineeship on Grant MH016745 from the National Institute of Mental Health, and Simon P. Livversedge was supported as a sabbatical visitor by British Academy Grant SG-35469.

We thank Chuck Clifton and three anonymous reviewers for their helpful comments on a draft of this article. Portions of the data were presented at the European Conference on Eye Movements, Dundee, Scotland, August 2003.

Correspondence concerning this article should be addressed to Keith Rayner, Department of Psychology, University of Massachusetts, Amherst, MA 01003. E-mail: rayner@psych.umass.edu

to determine the earliest point at which plausibility has an effect with respect to a specific target word.

Among studies that have examined eye movements and plausibility effects, there are a wide range of conclusions regarding how early in processing plausibility can have an effect on processing a word. At one extreme, Murray (1998; Murray & Rowan, 1998) has reported data suggesting that plausibility effects are apparent during parafoveal processing (a so-called *parafoveal-on-foveal* effect). In other words, Murray (1998) suggested that the implausibility of a certain target word can be detected before that word is actually fixated (i.e., when the implausible word is the word to the right of fixation). There is some uncertainty about the robustness of this result, as attempts at replication have proved difficult or elusive (Rayner & Juhasz, 2004; Rayner, White, Kambe, Miller, & Liversedge, 2003; though see Kennedy, Murray, & Boissiere, 2004, for a replication and acknowledgment of the difficulty of replicating the effects). At the other extreme, some studies (Boland & Blodgett, 2001; Garrod & Terras, 2000; Traxler, Pickering, & Clifton, 1998) have reported that effects of plausibility occur in the eye-movement record downstream from the target word at which they might first have been expected to occur. Such experiments suggest that there is a time delay between initially fixating an implausible word and the time at which that implausibility first affects fixation times. Finally, other studies have shown immediate effects of processing such that fixations on an implausible target word were longer than they were on a matched plausible target word (Cook & Myers, 2004). Thus, there is inconsistency among studies regarding the earliest influences of plausibility information on eye-movement behavior during reading.

Experiments by Braze, Shankweiler, Ni, and Palumbo (2002) and by Ni, Fodor, Crain, and Shankweiler (1998) in principle come close to addressing the question we set out to answer. Ni et al. recorded readers' eye movements as they read sentences like those in Sentence 1:

- 1a. *It seems that the cats won't usually /eat the/ food we put on the porch.*
- 1b. *It seems that the cats won't usually /bake the/ food we put on the porch.*
- 1c. *It seems that the cats won't usually /eating the/ food we put on the porch.*

Sentence 1a is a baseline control condition, whereas in Sentence 1b, there is what Ni et al. (1998) refer to as a pragmatic anomaly, and in Sentence 1c, there is what they refer to as a syntactic anomaly. Reading-time measures showed numerical differences on the critical region (which appears between slashes in Sentences 1a–1c), but none of these differences were significant across conditions (either in raw reading time or via a residual reading-time correction to adjust for length differences across conditions). These nonsignificant differences in first-pass reading times on the critical region included a 70-ms slowdown in raw reading times for the syntactic anomaly condition compared with the other two conditions, whereas residual times showed an entirely different pattern and were longest for the pragmatic anomaly conditions. These results suggest that there was considerable variability in their data, probably because of length differences in the critical region. Indeed, no reliable differences emerged in the reading-time

measures until the final region of the sentence (*on the porch*). However, there was one eye-movement measure, frequency of regressions, that did produce reliable differences at the critical region (and the two subsequent regions). The syntactic anomaly condition induced more regressions than the pragmatic anomaly condition, which in turn induced more regressions than the control condition, across the three regions.

Braze et al.'s (2002) study was quite similar to that of Ni et al. (1998). They recorded eye movements while readers read sentences like those in Sentence 2:

- 2a. *The wall will surely /crack after/ a few years in this harsh climate.*
- 2b. *The wall will surely /bite after/ a few years in this harsh climate.*
- 2c. *The wall will surely /cracking after/ a few years in this harsh climate.*

As in the Ni et al. (1998) study, Sentence 2a is a baseline condition, Sentence 2b is a pragmatic anomaly condition, and Sentence 2c is a syntactic anomaly condition. Braze et al. (2002) reported only the length-corrected first-pass reading time but did find an immediate effect in the region that we have placed between slashes in the example such that reading times were significantly longer for the two anomalous conditions than they were for the control. They also found that more regressions were launched from the target region in the syntactic anomaly condition than from the other two conditions.

Although the Ni et al. (1998) and Braze et al. (2002) experiments are related to our goal of determining how early plausibility influences fixation times, there are problems associated with them. In both experiments, the target region across conditions consisted of different words (which would add variability because lexical factors affect fixation times), and the words differed (sometimes considerably) in word length. Although length correction procedures are designed to deal with such differences, these procedures are much less reliable when the target region is short. Furthermore, careful examination of the Braze et al. materials indicates that the word that followed the manipulated word (*crack*, *bite*, *cracking*) varied considerably in length (i.e., it was as short as two letters and as long as seven letters). Given that word length has a very strong influence on the probability of fixation and overall reading time for a word (Rayner, 1998), there are clear problems of interpretation with these studies.

The research by Murray and colleagues (Kennedy et al., 2004; Murray, 1998; Murray & Rowan, 1998) is also highly relevant to the question we are asking. In these experiments, participants read a sentence displayed on a computer screen. They then pushed a button, and a second sentence appeared. Their task was to determine whether the two sentences were the same or different. Sentences like those in Version 3 were used:

- 3a. *The savages smacked the child.* (plausible–plausible, PP)
- 3b. *The savages smacked the money.* (plausible–implausible, PI)
- 3c. *The uranium smacked the child.* (implausible–plausible, IP)
- 3d. *The uranium smacked the money.* (implausible–implausible, II)

Sentence 3a is a control sentence, whereas in Sentences 3b–3d, various combinations of agents and verbs and/or verbs and themes are implausible. As we mentioned above, in these experiments,

there was some indication that readers were aware of the implausibility prior to fixating on the word that makes the implausibility apparent. Note that such a finding has profound implications for the relationship between the time course of oculomotor control processes (decisions about when and where to move the eyes during reading) and the time course of linguistic processing. That is, such parafoveal-on-foveal effects in the studies by Murray and colleagues (Kennedy et al., 2004; Murray, 1998; Murray & Rowan, 1998) would suggest that readers lexically identified a word, then processed it syntactically, and then semantically evaluated that text (at least at a shallow level), all prior to the word being directly fixated. Furthermore, the notion of parafoveal-on-foveal effects suggests that the meaning of two words is processed in parallel (see Starr & Rayner, 2001).

However, we have two concerns regarding this research. First, norming studies (see below) that we have done with their sentences raise some questions about the validity of their materials. For example, one of their sets of materials uses *hunters* as the first noun in the PP condition and *bishops* as the first noun in the IP condition when the rest of the sentence continues *stacked the bricks*. Our raters found either equally acceptable. Indeed, for 6 of their 24 sets of sentences, our raters rated the IP condition as or more acceptable than the PP condition. Furthermore, for another five of their sentences, our raters were not able to discriminate differences among the conditions (so that the PP, IP, PI, and II ratings were fairly equivalent). Our second concern is that an attempted replication (see Rayner, White, et al., 2003) that required participants to simply read the sentences, rather than to perform sentence matching, as per the studies conducted by Murray and colleagues (Kennedy et al., 2004; Murray, 1998; Murray & Rowan, 1998), failed to replicate the parafoveal-on-foveal effect. Indeed, there were effects when the implausible word was directly fixated, but there was no hint of an effect when readers were fixated on the preceding word.

Given these concerns about these studies, it seems necessary to further examine the influence of the implausibility of a critical target word on eye-movement behavior with a well controlled set of experimental stimuli. In the present experiment, we manipulated plausibility via the construction of sentences that had identical syntactic structure but that systematically differed in terms of whether a critical noun phrase (NP) contained in the sentences was an appropriate or inappropriate thematic-role filler for an accompanying verb. Furthermore, in the cases in which it was an appropriate role filler for the verb, we manipulated the degree to which it was appropriate in the context of the larger event. We selected NP arguments for the verbs in each of our three experimental conditions such that when thematic assignment occurred at the critical target word, the sentence was plausible (likely theme), implausible (unlikely theme given the implement used in the event), or anomalous<sup>1</sup> (inappropriate theme). Finally, it is critical to note that the target noun (and the determiner and adjective preceding it) was identical across the three conditions. Consider Sentences 4a–4c:

4a. *John used a knife to chop the large carrots for dinner.* (plausible control)

4b. *John used an axe to chop the large carrots for dinner.* (implausible)

4c. *John used a pump to inflate the large carrots for dinner.* (anomalous)

The experimental sentences started with a subject NP (e.g., either a proper noun like *John* in the example, or a definite description such as *The woman*). This was followed by the finite transitive verb *used*, after which came an NP that introduced an implement used in the event described by the sentence (e.g., *a knife*) and an infinitival verb (e.g., *to chop*). Finally, across all conditions there always followed the same adjectival NP (e.g., *the large carrots*) that was the direct object of the infinitival verb and was assigned the thematic role of theme. It is important to note that the adjective was chosen so that it would not modulate the plausibility of the direct-object NP.

The materials were constructed such that in all three conditions, the NP object of *used* was a likely implement to be used in the event described by the infinitival verb. Thus, *axe* and *knife* are plausible implements to be used in a *chopping event*, and *pump* is a plausible implement in an *inflating event*. Furthermore, in the plausible (Sentence 4a) and implausible (Sentence 4b) versions, the critical NP was always a plausible theme of the infinitival verb (i.e., carrots are a plausible thing to chop). In the plausible control condition (Sentence 4a), the implement (*knife*) is natural to use in the event described by the infinitival verb and its theme (*chopping carrots*), so when the target word is read, assignment of thematic roles and integration into an event schema results in an interpretation of the sentence that is semantically congruous.

In contrast, in the implausible version (Sentence 4b), the implement (*axe*) is plausible with the denotation of the verb (*to chop*), and the verb and its theme (*carrots*) are plausible together. However, when the target word is read and integrated into a semantic representation including both the implement and the verb, the result depicts an implausible event. To be clear, although it is plausible to perform the action of chopping with an axe, and carrots are a plausible entity to be chopped, it is implausible to perform the action of chopping with an axe when the entities being chopped are carrots. Thus, after the target word is assigned a thematic role and integrated into a semantic representation, detection of the implausibility associated with Sentence 4b should occur.

Finally, in the anomalous condition (Sentence 4c), although the implement used in the event (*pump*) is consistent with the type of event (*inflate*), the adjectival NP (*carrots*) could not plausibly be assigned the role of theme by the verb. That is to say, under normal circumstances, carrots cannot be inflated. Thus, for the anomalous sentences, when the reader processes the target word and attempts to assign the role of theme to the direct-object NP, an anomaly should be apparent.

<sup>1</sup> We use the term *anomalous* largely as a matter of convenience. We fully realize that there may well be scenarios under which the events described in our so-called anomalous conditions might occur. For example, in a cartoon situation, one might be able to envision a pump used to inflate a plastic carrot. However, in the general course of events, the anomalous conditions reflect extremely unlikely (if not impossible) real-world events. Furthermore, the normative data we collected (see below) provide some validation of our classifications.

In the current experiment, we examined the time course of plausibility effects with respect to the target word (*carrots*). We examined a number of standard eye-movement reading-time measures (see Liversedge & Findlay, 2000; Rayner, 1998). We anticipated that we would observe the earliest disruption to processing for the target word in the anomalous condition compared with the control (plausible) condition. Thus, we expected that the anomalous condition would yield disruption effects in first-pass reading time on the target word. Given Murray's results (Kennedy et al., 2004; Murray, 1998; Murray & Rowan, 1998), we also examined processing on the word preceding the target word. With respect to the implausible condition, we expected that the severity of the disruption would be reduced in magnitude and would be less immediate, appearing later in the eye-movement record.

## Method

### Participants

Thirty-six undergraduate students at the University of Massachusetts, Amherst with normal or corrected-to-normal vision (with soft contact lenses) participated in the study<sup>2</sup>. They were all naive concerning the purpose of the experiment; they received either course credit or a nominal sum (\$8) in payment.

### Apparatus

A Fourward Technologies Dual-Purkinje Image Generation VI eye-tracker monitored gaze location and participants' right eyes during reading (although viewing was binocular). The eye-tracker has a spatial resolution of 10-min arc (so we know precisely where readers were fixated). Materials were displayed on a personal computer on a monitor 61 cm from participants' eyes. Gaze location was monitored every millisecond to produce a sequence of fixations with start and finish times.

### Materials

We constructed 30 experimental items, each with a plausible form (such as Sentence 4a in the example), an implausible form (Sentence 4b), and an anomalous form (Sentence 4c). We designed items so that the plausibility violation always occurred at the noun of the adjectival NP (the critical target word) following the infinitival verb. All of the words following the infinitival verb were the same across conditions. The critical word was always a minimum of five characters long to increase the likelihood that readers would fixate it. We selected the adjective that preceded the noun carefully so as not to modulate the plausibility of the critical noun as the object of the verb.

Norming data obtained from 12 University of Massachusetts, Amherst students, who did not participate in the main experiment, provided validation of our classification. These participants were provided with one version of each sentence and asked to rate (on a 1–5 scale) whether the sentence was normal (a rating of 5), highly unlikely to occur in the real world (a rating of 1) or unlikely to occur but possible (a rating of 3). Participants were encouraged to use the full range of the scale. The words *anomalous* and *implausible* were not included in the instructions for rating. The versions of the sentence that we categorized as anomalous received a mean rating of 1.4, those we categorized as implausible received a mean rating of 2.9, and the control sentences received a rating of 4.8. It is critical to note that for every single sentence, the ratings coincided with our categorization so that the rating for the control (plausible) condition was rated highest, the implausible was next, and the anomalous was rated lowest<sup>3</sup>. A complete list of the sentences and the ratings values associated with each version is provided in the Appendix.

We combined the 30 experimental items with 90 filler sentences. Five of the filler sentences were implausible or anomalous. An additional 10 filler sentences began with a proper name and the verb *used* but then continued with a syntactic structure different from that of the experimental items. We included these fillers to avoid predictability of syntactic structure within the experimental materials. We constructed three files of materials such that 10 items appeared in each condition in each file and each item appeared in a different condition in each file. After 25% of the sentences, participants were required to answer a yes/no comprehension question (half of which required a "yes" response). Questions were only asked about the experimental items in the plausible condition.

### Procedure

Experimental sessions lasted between 30 and 45 min. We minimized head movements using forehead restraints and a bite bar. Participants were instructed that the sentences they read would vary in length and difficulty and that some of them might be "a little weird." They were asked to read normally, for comprehension, and were told that after 25% of the sentences, they would be required to respond to a comprehension question by pressing a *yes* or a *no* button. Once the participant was seated at the eye-tracker, it was aligned and calibrated. After each sentence, a pattern of boxes appeared on the screen as a calibration check. The eye-tracker was recalibrated when a dot that moved in synchrony with the participants' eye movements did not appear within the calibration boxes.

## Results and Discussion

We divided the target sentences into three regions of interest for analysis. These regions were (a) the pretarget region made up of determiner and adjective (*the large*), (b) the target-word region (*carrots*), and (c) the posttarget region (*for dinner*). We subjected the data for each region to analyses of variance (ANOVAs), one for participants ( $F_1$ ) and the other for items ( $F_2$ ). Comprehension was high (all participants scored 80% or better in response to the questions). Less than 1% of the data were lost because of track losses. The target words were fixated 95% of the time during first-pass reading in each of the three conditions.

The eye-movement record yields a number of measures that are associated with variations in the time course of processing a target word (Liversedge & Findlay, 2000; Rayner, 1998): first-fixation duration, single-fixation duration, gaze duration, go-past time, and total reading time. The first four measures all reflect first-pass reading time, whereas the last measure includes regressions to the target word (and hence also involves second-pass reading). Each successive measure reflects later processing activities. The first two measures reflect the reader's initial encounter with a target word: First-fixation duration is the duration of the first fixation on a word independent of the number of first-pass fixations (and thus

<sup>2</sup> In addition to the 36 participants whose data are reported, we excluded data from 2 others because they had comprehension rates less than 80%.

<sup>3</sup> The sentence sets used by Murray (1998; Murray & Rowan, 1998) and Kennedy et al. (2004) were also rated by these same participants. The sentences were corrected to account for British versus American spellings. The mean ratings (on the same scale) for their sentences were as follows: PP = 3.7, PI = 2.0, IP = 2.6, and II = 1.9.



Table 1  
*Fixation Times on the Pretarget Region (in ms)*

Condition	FFD		Single		Gaze		Last 3		Go-past		Total	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
All items												
Anomalous	255	34	275	43	348	65	261	68	373	75	453	150
Implausible	243	32	270	45	334	65	231	49	352	67	369	66
Control	248	32	270	40	331	66	241	35	345	68	359	74
Supplemental analysis (18 items)												
Anomalous	254	39	269	44	356	73	254	59	384	83	457	157
Implausible	241	36	267	43	333	73	227	63	360	78	371	72
Control	249	36	264	46	338	81	228	51	349	84	370	86

*Note.* FFD = first-fixation duration.

can be either a single fixation or the first of multiple fixations), whereas single-fixation duration is the duration of the fixation on a word in those cases in which there is only one fixation on the word. Gaze duration is the sum of all fixations on a target word before the eyes leave the word and thus includes the initial fixation time on the target word as well as any refixations on the word (prior to moving to another word). The go-past measure includes the amount of time that the reader looks at the target word as well as any time spent rereading earlier parts of the sentence before moving ahead to inspect new portions of the sentence. It most likely reflects both lexical processing and integration processes, because the reader likely realized that there was some problem with the target word and thus made a regression back to some earlier part of the sentence. Given that Ni et al. (1998) and Braze et al. (2002) found immediate effects on regressions, both go-past and the frequency of regressions are reported here.

As noted previously, our hypothesis was that the disruption due to an anomalous target word would appear earlier in the eye-movement record and be greater than the disruption that was due to an implausible word. We first discuss the first-pass reading data for the pretarget region, followed by the target region, the post-target region, and then total reading time. Given Murray's results (Kennedy et al., 2004; Murray, 1998; Murray & Rowan, 1998), we devote more attention to the pretarget region than to the other regions (where the results are straightforward). In all regions, fixations under 80 ms were combined with fixations on adjacent letters, whereas fixations under 80 ms and not within three letters of another fixation were eliminated. In addition, all data points more than 2.5 standard deviations from the mean value were eliminated (resulting in a loss of less than 3% of the data). Altogether, less than 10% of the data were lost because of track losses, target-word skipping, and data trimming.

### Pretarget Region

Table 1 shows the first-fixation duration, single-fixation duration, and gaze duration data for the pretarget region. ANOVAs comparing the three conditions showed no significant differences for the first-fixation and single-fixation data<sup>4</sup>. Furthermore, none of the gaze duration differences were close to significant. How-

ever, the anomalous condition yielded gaze durations that were 14 ms and 17 ms longer than the control and implausible conditions, respectively. In order to more completely determine whether there was any effect in the pretarget region (and following Murray, 1998), we computed the average fixation duration when the eyes were still in the pretarget region but fixated three or fewer letters away from the beginning of the target word<sup>5</sup>. When this was done, there was a further hint of an effect (see Table 1), as fixations in the anomalous condition averaged 261 ms compared with 236 ms in the implausible and control conditions (which did not differ from each other,  $t < 1$ ). The difference between the anomalous condition and the other two conditions was significant,  $t_1(32) = 2.32, p < .05$ ;  $t_2(24) = 2.55, p < .05$ .

Is this difference between the anomalous condition and the other two conditions evidence of a parafoveal-on-foveal effect (in which readers are concurrently lexically processing both the fixated word and the word to its right)? It is possible that it is, but we think it is more likely that the effect is due to the reader processing word

<sup>4</sup> Actually, for the participants analysis for first-fixation duration, there was a significant difference among the conditions,  $F_1(2, 70) = 3.93, p < .05$ ;  $F_2(2, 58) = 1.45, p > .24$ . The difference was that the anomalous condition was longer than the implausible condition,  $t_1(35) = 2.50, p < .05$ . However, this effect was entirely due to first fixations that initially landed quite close to the target region (thus on the last three letters of the pretarget region). When these fixations were eliminated from the analysis, the means for the three conditions were 247 ms, 240 ms, and 244 ms (for the anomalous, implausible, and control conditions, respectively), and there was no difference among them ( $ps > .25$ ).

<sup>5</sup> Trials with regressions from the pretarget region were not included because they could potentially include cases where readers had a parafoveal preview of the target word, regressed, continued reading, and fixated on one of the three characters before the target region before fixating the target word. In addition, cases in which readers skipped the target word on the ensuing saccade were not included in this analysis (the frequency with which the target word was skipped on saccades launched from this three-character region did not differ across conditions). Finally, the degrees of freedom for both the participants and items analyses are different from other analyses reported here because of missing data for some participants and items.

$n$  while actually fixated on word  $n - 1$ . There are three ways in which this could happen; none of these three alternative explanations are mutually exclusive, and none of them requires that parallel lexical processing of the fixated word and the word adjacent to it must occur. First, although it occurs much less frequently (if at all) during reading than nonreading attention tasks (Rayner, 1998), it is possible for the locus of attention to differ from the eye position. Second, if one's eye-tracking system was not highly accurate or well calibrated, it is possible for the reader to be actually processing word  $n$  while the eye-tracker indicates that he or she is actually fixating word  $n - 1$ <sup>6</sup>. Third, it is well known that during reading (McConkie, Kerr, Reddix, & Zola, 1988; Rayner, 1998) there are saccadic undershoots wherein the reader intends to make a saccade to a target word (word  $n$ ), but the actual eye movement falls short of the target (resulting in a fixation on word  $n - 1$ ). It is, of course, impossible to tell whether, for any particular saccade, a reader has undershot an intended target word. However, examination of the three-letter region in question revealed that 22% of all fixations in the pretarget region were in this region. If we assume, relatively conservatively, that about a third of those fixations were due to undershoots, then those fixations would reflect processing associated with the target word rather than the word on which the eye was currently fixated (word  $n - 1$ ), and any such fixation duration differences would actually be due to the reader processing the target word.

Examination of the go-past measure also suggested that the anomalous condition was processed differently than the other two conditions, as there was an effect of condition,  $F_1(2, 70) = 7.10$ ,  $p < .01$ ;  $F_2(2, 58) = 3.31$ ,  $p < .05$ , and pairwise  $t$  tests indicated that the implausible and control conditions did not differ from each other ( $ps > .10$ ), but the anomalous condition differed from the other two conditions,  $t_1(35) = 3.43$ ,  $p < .01$ ;  $t_2(29) = 2.00$ ,  $p = .055$ . Two points are relevant with respect to this finding. First, all data from the pretarget region were included in this analysis (including fixations in the last three characters of this region). Second, it is instructive that the measure that Murray and colleagues used in their studies (see Kennedy et al., 2004; Murray, 1998; Murray & Rowan, 1998) is in fact equivalent to the go-past measure. That is, they did not report the more typical measures (see Liversedge & Findlay, 2000; Rayner, 1998) that are reported in our tables but, rather, the go-past measure was used as their measure of first-pass reading time. Although there were no significant differences in the frequency of regressions out of the pretarget region (6.3% for anomalous, 4.4% for implausible, and 3.9% for control;  $ps > .35$ ), it is interesting that for the three-letter region in which we found an effect of anomaly on fixation duration, over half of the regressions from that region were in the anomalous condition. The fact that there were more regressions out of this region in the anomalous condition would lead to the longer go-past times in the region. Our finding that there is a difference in go-past time is therefore also consistent with the explanation that we provided above as the basis for the effect in the last three-character region of the pretarget region. To be specific, we suspect that the effect of the anomalous word in the three-character region preceding the target word and the go-past results are both most likely due to a mismatch in actual eye location and the word being processed and not due to parafoveal-on-foveal processing

wherein the meaning of the word to the right of fixation is influencing the current fixation.

Before moving to the results for the target-word region, it is important to consider an alternative explanation for the effects that occurred in the pretarget region. To be specific, although the experimental sentences were designed so that the plausibility violation always occurred at the noun of the adjectival NP (and the norming study reported above confirmed the validity of the designation), perhaps there were uncontrolled differences in how well the adjective itself fit with what had been read up to the point that it was encountered. That is, if the adjective did not fit particularly well in the anomalous condition, longer fixations would be expected, and these longer fixations would be due to the adjective and not to the noun<sup>7</sup>. In order to evaluate this possibility, we had two groups of participants rate how well the adjective fit with the preceding sentence. One group of 10 raters was given the experimental sentences up to the adjective and asked to rate how well the adjective fit with what had come before (a 5-point rating scale was used, with 5 indicating that the adjective fit very well and 1 indicating that it did not fit at all). For this group of raters, there were no fillers to act as anchor points. Another group of 7 raters were given the same task, but they also received 15 fillers with very anomalous adjectives. In fact, the ratings came out the same for both groups, so we will discuss only those with no fillers.

It is interesting to note that there were significant differences across the conditions,  $F(2, 58) = 10.19$ ,  $p < .01$ , in terms of how well the adjective fit with the prior sentence frame; the adjective in the anomalous (4.1) and implausible (3.9) conditions did not fit as well as in the control condition (4.4). Note though that the nature of the difference in the norming study was that the anomalous and implausible conditions differed from the control condition, whereas in the eye fixation times, the anomalous differed from the implausible and control conditions. More important, however, a supplementary analysis of the data shown in Table 1 based on 18 sentences (those marked with an asterisk in the Appendix) in which there was no difference ( $p > .28$ ) across the conditions (with mean ratings of 4.2, 4.1, and 4.2 for the anomalous, implausible, and control conditions, respectively) yielded exactly the same pattern of data for the pretarget region (see Table 1). Furthermore, all significant effects for the other regions (for all measures) remained significant with these supplementary analyses; the mean values for the supplementary analyses are presented in Tables 1–3.

### Target Word

Table 2 shows the fixation times on the target word. Although there were no differences among the three conditions for first-

<sup>6</sup> Although the Dual-Purkinje tracker that we used has high spatial resolution, and although we checked the calibration on each trial, it is still possible that small head movements would result in a less-than-perfect calibration on any given trial. Of course, less accurate eye-tracking systems would have this problem more frequently.

<sup>7</sup> We thank an anonymous reviewer and Maryellen MacDonald for pointing out this possibility.

Table 2  
Fixation Times on the Target Word (in ms)

Condition	FFD		Single		Gaze		Go-past		Total	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
All items										
Anomalous	268	40	277	46	306	53	332	73	390	121
Implausible	261	35	270	39	289	39	310	37	312	48
Control	262	38	265	46	286	49	297	57	297	52
Supplemental analysis (18 items)										
Anomalous	266	47	280	54	308	65	339	85	383	112
Implausible	262	36	268	38	286	37	307	41	312	44
Control	261	41	263	48	282	46	295	60	291	48

Note. FFD = first-fixation duration.

fixation duration and single-fixation duration<sup>8</sup> (all  $ps > .15$ ), there was a significant difference in gaze duration,  $F_1(2, 70) = 3.44$ ,  $p < .05$ ; and  $F_2(2, 58) = 4.50$ ,  $p < .05$ . It is more critical to note that pairwise  $t$  tests revealed significant differences between the anomalous and control conditions,  $t_1(35) = 2.47$ ,  $p < .05$ ;  $t_2(29) = 2.24$ ,  $p < .05$ , and between the anomalous and implausible conditions,  $t_1(35) = 2.03$ ,  $p < .05$ ;  $t_2(29) = 2.57$ ,  $p < .05$ . The implausible and control conditions did not differ ( $ts < 1$ ).

The go-past measure also yielded a significant main effect,  $F_1(2, 70) = 6.45$ ,  $p < .01$ ; and  $F_2(2, 58) = 7.82$ ,  $p < .01$ . Pairwise comparisons revealed a separation between the times for the different conditions: anomalous versus control,  $t_1(35) = 3.08$ ,  $p < .01$ ;  $t_2(29) = 3.63$ ,  $p < .01$ ; anomalous versus implausible,  $t_1(35) = 2.14$ ,  $p < .05$ ;  $t_2(29) = 2.32$ ,  $p < .05$ ; and implausible versus control,  $t_1(35) = 1.75$ ,  $p = .089$ ;  $t_2(29) = 1.71$ ,  $p = .098$ . We note the latter difference because it represents the first hint of a difference between these two conditions.

### Posttarget Region

Table 3 shows the reading-time measures for the posttarget region. There were differences across all four measures: first-

Table 3  
Fixation Times on the Posttarget Region (in ms)

Condition	FFD		Single		Gaze		Go-past		Total	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
All items										
Anomalous	288	58	299	75	351	84	493	226	416	112
Implausible	261	39	261	45	325	75	362	88	354	85
Control	267	48	268	54	312	64	325	76	325	65
Supplementary analysis (18 items)										
Anomalous	293	75	291	79	344	94	434	183	398	121
Implausible	255	46	259	51	303	88	341	107	344	105
Control	258	52	266	61	311	84	325	98	320	79

Note. FFD = first-fixation duration.

fixation duration,  $F_1(2, 70) = 7.89$ ,  $p < .01$ ;  $F_2(2, 58) = 5.32$ ,  $p < .01$ ; single-fixation duration,  $F_1(2, 70) = 5.17$ ,  $p < .01$ ;  $F_2(2, 58) = 6.25$ ,  $p < .01$ ; gaze duration,  $F_1(2, 70) = 5.78$ ,  $p < .01$ ;  $F_2(2, 58) = 3.97$ ,  $p < .05$ ; and go-past,  $F_1(2, 70) = 27.74$ ,  $p < .01$ ;  $F_2(2, 58) = 17.8$ ,  $p < .01$ . Pairwise comparisons indicated that the basic pattern for the first three measures was that the anomalous condition differed (all  $ps < .05$ ) from the implausible and control conditions, but they did not differ from each other ( $ps > .18$ ). In contrast, the go-past measure again indicated a separation of the three conditions: anomalous versus control,  $t_1(35) = 6.59$ ,  $p < .01$ ;  $t_2(29) = 6.37$ ,  $p < .01$ ; anomalous versus implausible,  $t_1(35) = 4.77$ ,  $p < .01$ ;  $t_2(29) = 3.24$ ,  $p < .01$ ; and implausible versus control,  $t_1(35) = 2.76$ ,  $p < .01$ ;  $t_2(29) = 2.36$ ,  $p < .05$ . The probability of regressing out of the posttarget region was 3% for the control condition, 6% for the implausible condition, and 23% for the anomalous condition,  $F_1(2, 70) = 19.71$ ,  $p < .01$ ;  $F_2(2, 58) = 16.62$ ,  $p < .01$ , and the anomalous condition differed from each of the other two conditions ( $ps < .01$ ), which did not differ significantly from each other.

### Total Times

Total reading-time data are included in each table. There were differences across conditions in all regions (all  $ps < .01$ ). In the pretarget region, the anomalous condition differed ( $ps < .01$ ) from the other two conditions, which did not differ from each other ( $ps > .25$ ). For the target word, the anomalous condition again differed ( $ps < .01$ ) from the other two conditions, with the difference between the implausible and control condition being significant with  $t_2(29) = 2.06$ ,  $p < .05$ , but not  $t_1(35) = 1.63$ ,  $p =$

<sup>8</sup> It appears on the surface that there is some inconsistency between our finding of an anomaly effect when the eyes were on the last three letters of the pretarget region but no effect on single-fixation durations in the target region. However, close examination of the data revealed that when there was no trimming of data points more than 2.5 standard deviations from the mean, there was a difference in single-fixation duration between the anomalous condition (288 ms) and the mean of the other two conditions (272 ms and 270 ms for the implausible and control conditions, respectively),  $t_1(35) = 2.15$ ,  $p < .05$ ;  $t_2(29) = 2.26$ ,  $p < .05$ .

.11. Finally, in the posttarget region, all differences were significant ( $ps < .05$ ).

### General Discussion

The experimental manipulations used in the experiment affected fixation/reading times. With respect to the issue of how early in processing plausibility effects appear, our results are clear. The effect on an anomalous word was evident on the gaze duration on the word. Interestingly, there was no effect on first-fixation duration or single-fixation duration, so the difference on gaze duration is due to the fact that readers were more likely to refixate on the anomalous word before moving their eyes to another word. When we examined the go-past measure, we did find some evidence that there was an effect of implausibility (as well as an effect due to anomaly).

When we considered the pretarget region, we did find evidence to indicate that there was an effect of the anomalous target word when the reader was fixated just to the left of it. Does this reflect a parafoveal-on-foveal effect? It is possible that it does, but we have suggested that it is also at least as likely that the effect is due to the reader processing word  $n$  while actually fixated on word  $n - 1$ . Though we outlined three possible reasons for why this might occur, we suspect that saccadic undershoots count for much of the effect. That is, saccades in reading sometimes undershoot the intended target location (McConkie et al., 1988; Rayner, 1998) so that after the saccade, the reader ends up fixating on word  $n - 1$  even though he or she is actually processing word  $n$ . We suspect that saccadic undershoots are at least as likely an explanation of the finding as is parafoveal-on-foveal processing. Furthermore, because Murray (1998) and Kennedy et al. (2004) used the go-past measure as their index of first-pass processing time, we again suspect that the same type of explanation holds for their results. That is, if the reader was really processing the target word on a small percentage of fixations (while actually fixated on word  $n - 1$ ), it would not be at all surprising that a regression would be initiated (in response to the strangeness of word  $n$ ) in such a situation (therefore inflating the go-past time).

When the reader was fixated directly on the target word, there was clear disruption due to anomaly evident in gaze duration. Thus, the effect of anomaly "hit the reader over the head"; the effect was immediate. The effect of implausibility, on the other hand, was not nearly as immediate. There was no effect of implausibility on gaze duration. However, the reader's decision to move forward versus backward in the text was influenced by implausibility (as evidenced in the go-past measure), but the effect was weaker and more short-lived than the influence of anomaly.

It is also interesting to note that in the posttarget region (the region that appears toward the end of the sentence), readers were far more likely to make a regression in order to reinspect earlier portions of the sentence when it was anomalous compared with when it was implausible or plausible (in the control condition). This suggests that attempts to form a coherent semantic representation of an anomalous sentence required readers, after initially reading through the sentence, to spend considerable time rereading it, whereas for both implausible and control sentences, rereading was required to a far lesser extent. Thus anomalous violations, unlike implausible violations, served as primary factors in trigger-

ing regressive saccades to permit rereading in an attempt to derive a meaningful interpretation of the sentence.

The results of this study suggest that the timing of plausibility violation detection may be related to the violation's severity. If this is the case, then the differences that previous studies have found in the timing of plausibility violation detection may be related to variation in the severity of the violations used. For example, it is possible that Ni et al. (1998) found only late eye-movement effects of plausibility violations because they tested relatively less severe violations, whereas Braze et al. (2002) found evidence of disruption on the word after the anomaly because they tested more severe violations. Unfortunately, the relative severity of violations between and within these experiments was not controlled, so this is speculative.

Although the severity of the plausibility violations was manipulated in our experiment, the implausible and anomalous conditions also differed with respect to whether the violation occurred in a theta-assigning relation. In the implausible conditions, the violation arose between two objects participating in an event, whereas in the anomalous conditions, for the majority of our materials the violation arose in the relation between a theta-assigning verb and one of its arguments. This raises the possibility that violations in the anomalous conditions were detected quickly because, in most cases, they could be detected on the basis of purely lexical information, assuming that information associated with a verb's lexical entry may serve to license certain nouns as verb arguments but not others. In the implausible conditions, violations may have been detected more slowly because they may have arisen at a stage of processing after theta assignment, when the target word was integrated into a semantic representation of the sentence fragment up to that point. Our results are therefore consistent with the suggestion that qualitatively different types of processing take place at different stages during sentence comprehension.

The results of our study also have implications for the use of plausibility manipulations to disambiguate potential syntactic ambiguities in garden-path sentences. The finding that gaze durations on the target word were affected when an anomalous word was read but not when an implausible word was read indicates that early stages of language comprehension are sensitive only to severe plausibility violations. Less severe violations that arise in the acceptability of the relationship between two objects in an event show delayed effects of a smaller magnitude in the eye-movement record. These smaller delayed effects presumably influence only later stages of processing during comprehension. Our findings suggest that in studies examining aspects of syntactic processing, any plausibility manipulations intended to disambiguate syntactic ambiguities during the initial stage of construction must result in a severe anomaly, as those are the only violations that affect the very early stages of processing reflected in the eye-movement record. Less extreme plausibility violations are not detected until later stages of processing and thus might not be appropriate to use if the experimenter wishes to determine the earliest point during comprehension at which a misanalysis was first detected.

This conclusion seems to conflict with results from experiments using head-mounted eye-tracking and the visual-world paradigm that suggest extremely early plausibility effects during the interpretation of auditorially presented sentences (Altmann & Kamide,



1999; Kamide, Altmann & Haywood, 2003). These effects take the form of early predictions about the likely identity of a direct object based on the subject and verb. Although these effects suggest that plausibility informs early stages of language comprehension, it is possible that they arise because the domain of discourse has been artificially restricted to an extremely small set of objects, only one of which is a plausible direct-object candidate. In normal reading, no such restriction applies, and therefore, arguments are not usually this predictable. Consequently, readers should be less likely to anticipate arguments on the basis of plausibility, and therefore, it is not surprising to find later effects of plausibility in reading studies than in visual-world studies.

It is also interesting to consider the extent to which our results reflect mere transitional probabilities between words. That is, McDonald and Shillcock (2003a, 2003b) recently showed that the frequency with which two words cooccur influences the amount of time that readers look at the second word in a pair of words (even when predictability is very low). Thus, readers look longer at *losses* following *accept* than at *defeat* following *accept* because the former pair cooccurs less frequently than the latter pair. Thus, it could be argued that gaze durations are longer on our target noun (*carrots*) in the anomalous case because *chop carrots* occurs more frequently than *inflate carrots* (in fact, the latter two words probably never cooccur; see Footnote 1). McDonald and Shillcock's finding is interesting, but it must be noted that the effect was rather small (15–20 ms in gaze duration) and primarily occurred when the prior fixation was close to the beginning of the second word. It is therefore unlikely that the more substantive effects we obtained can be explained (at least in their entirety) by transitional probabilities between words. Also, we suspect that it is somewhat unlikely that transitional probability effects would survive intervening words (and we always had a determiner and an adjective between the verb and target noun). Just as priming effects on eye fixations do not extend across a series of words and major syntactic breaks (Carroll & Slowiaczek, 1986; Morris, 1994), we suspect that such transitional probability effects do not survive across intervening words.

Finally, with respect to the issues we raised at the outset regarding whether plausibility values should be incorporated into models of eye-movement control, if the goal is to capture the full range of variables that have an influence on fixation times, then adding plausibility may indeed improve the fits between model and data. However, if the goal is to describe the most typical situation in which the eyes move forward in the text, it may not be necessary to do so. Apparently, only in cases of an extreme plausibility violation is the normal lexical processing that drives the eyes through the text (Engbert et al., 2002; Liversedge et al., 2004; Rayner, Liversedge, White, & Vergilino-Perez, 2003; Reichle et al., 1998, 2003) short-circuited or interrupted (with the decision to move the eyes delayed). Less serious plausibility violations apparently affect only later stages in the processing sequence.

## References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of sentence reference. *Cognition*, *73*, 247–264.
- Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, *45*, 391–411.
- Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research*, *31*, 25–44.
- Carroll, P. J., & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition*, *14*, 509–522.
- Clifton, C., Jr. (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology*, *47*, 222–246.
- Clifton, C., Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317–334.
- Cook, A., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influences of context and world knowledge. *Journal of Memory and Language*, *50*, 268–288.
- 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.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.
- Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language*, *42*, 526–544.
- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*, 431–439.
- Juhász, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye-fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1312–1318.
- Juhász, B. J., & Rayner, K. (in press). The role of age-of-acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156.
- Kennedy, A., Murray, W. S., & Boissiere, C. (2004). Parafoveal pragmatics revisited. *European Journal of Cognitive Psychology*, *16*, 128–153.
- Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science*, *4*, 6–14.
- Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., & Kentridge, R. (2004). Eye movements when reading disappearing text: Is there a gap effect in reading? *Vision Research*, *44*, 1013–1024.
- 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.
- McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities. *Psychological Science*, *14*, 648–652.
- McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, *43*, 1735–1751.
- Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 92–103.
- Murray, W. S. (1998). Parafoveal pragmatics. In G. Underwood (Ed.), *Eye*

- guidance in reading and scene perception* (pp. 181–200). Oxford, England: Elsevier.
- Murray, W. S., & Rowan, M. (1998). Early, mandatory, pragmatic processing. *Journal of Psycholinguistic Research*, 27, 1–22.
- Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: Assessing the contributions of syntax, semantics, and plausibility in resolving ambiguities. *Language and Cognitive Processes*, 11, 283–334.
- Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research*, 27, 515–540.
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 940–961.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K., Carlson, M. A., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358–374.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191–201.
- Rayner, K., & Juhasz, B. J. (2004). Eye movements in reading: Old questions and new directions. *European Journal of Cognitive Psychology*, 16, 340–352.
- Rayner, K., Liveredge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. *Psychological Science*, 14, 385–389.
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504–509.
- Rayner, K., White, S. J., Kambe, G., Miller, B., & Liveredge, S. P. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 213–234). Oxford, England: Elsevier.
- 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, 445–476.
- Starr, M. S., & Rayner, K. (2001). Eye movements during reading: Some current controversies. *Trends in Cognitive Science*, 5, 156–163.
- Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, 48, 740–759.
- Traxler, M. J., Pickering, M. J., & Clifton, C., Jr. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39, 558–592.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318.

(Appendix follows)

## Appendix

## Sentences Used in the Experiment and Their Scores in Two Norming Studies

Sentences	Overall plausibility rating	Adjective fit
John used a pump to inflate the large carrots for dinner last night.	1.0	4.4*
John used an axe to chop the large carrots for dinner last night.	3.0	4.7
John used a knife to chop the large carrots for dinner last night.	5.0	4.8
The man used a grill to cook the steaming asphalt on the road.	1.5	3.8*
The man used a knife to spread the steaming asphalt on the road.	2.0	3.3
The man used a shovel to spread the steaming asphalt on the road.	4.8	3.3
The man used a feather to tickle the thin spaghetti yesterday evening.	1.0	3.6*
The man used a net to drain the thin spaghetti yesterday evening.	3.5	3.3
The man used a strainer to drain the thin spaghetti yesterday evening.	5.0	3.3
The woman used a book to teach the tough bread very carefully.	1.3	4.2*
The woman used a saw to cut the tough bread very carefully.	3.0	4.4
The woman used a knife to cut the tough bread very carefully.	4.5	4.6
The woman used the map to direct the large present yesterday.	1.3	4.0
The woman used the rags to wrap the large present yesterday.	3.5	3.4
The woman used the paper to wrap the large present yesterday.	5.0	4.5
Bill used the calculator to compute the hard cheese from Italy.	1.0	4.8*
Bill used the scissors to cut the hard cheese from Italy.	2.8	4.0
Bill used the knife to cut the hard cheese from Italy.	4.3	4.2
The man used an iron to flatten the stiff patient after surgery.	1.0	3.7
The man used a bribe to relax the stiff patient after surgery.	2.0	2.9
The man used a tranquilizer to relax the stiff patient after surgery.	4.0	4.3
The woman used a dagger to stab the dirty miniatures sitting on the shelf.	1.5	3.6
The woman used a vacuum to clean the dirty miniatures sitting on the shelf.	3.0	4.8
The woman used a duster to clean the dirty miniatures sitting on the shelf.	4.8	4.8
The man used a tank to ambush the front porch for the party.	1.8	3.8
The man used a toothbrush to clean the front porch for the party.	3.3	3.6
The man used a mop to clean the front porch for the party.	5.0	4.7
The hostess used the music to calm the hot beans for dinner.	1.3	3.4
The hostess used the toothpick to serve the hot beans for dinner.	3.3	3.5
The hostess used the dish to serve the hot beans for dinner.	5.0	4.6
The man used the phone to call the old frame together.	1.0	4.5
The man used the chopsticks to hold the old frame together.	3.3	3.2
The man used the clamp to hold the old frame together.	4.8	4.1
Jenny used a hose to water the small butterfly flying by.	2.0	4.6*
Jenny used a mousetrap to catch the small butterfly flying by.	3.0	4.9
Jenny used a net to catch the small butterfly flying by.	5.0	4.7
Patricia used a fork to eat the fresh water extremely carefully.	1.5	5.0
Patricia used a purse to carry the fresh water extremely carefully.	2.5	3.6
Patricia used a bucket to carry the fresh water extremely carefully.	4.3	4.7
Matthew used a broom to sweep the bright comet as it passed by.	1.0	3.4
Matthew used a microscope to watch the bright comet as it passed by.	3.5	3.1
Matthew used a telescope to watch the bright comet as it passed by.	5.0	4.7
George used a harness to restrain the many flowers in his garden.	2.5	4.1*
George used a sword to protect the many flowers in his garden.	2.8	4.5
George used a fence to protect the many flowers in his garden.	4.5	4.1
Frank used a hammer to nail the heavy groceries from the store.	1.8	4.5*
Frank used a helicopter to transport the heavy groceries from the store.	3.0	4.3
Frank used a cart to transport the heavy groceries from the store.	4.5	4.7

## Appendix

*Continued*

Sentences	Overall plausibility rating	Adjective fit
Julie used a needle to sew the various children after recess.	1.3	4.4*
Julie used a flare to summon the various children after recess.	3.5	4.0
Julie used a bell to summon the various children after recess.	5.0	4.4
Melinda used a spatula to flip the little house at night.	1.0	4.3*
Melinda used an anchor to secure the little house at night.	1.8	3.8
Melinda used a lock to secure the little house at night.	4.8	4.5
Donald used a jeep to climb the highest spire of the cathedral.	1.5	4.1*
Donald used a stool to reach the highest spire of the cathedral.	2.5	4.5
Donald used a ladder to reach the highest spire of the cathedral.	4.0	5.0
The man used a rag to polish the precious liquid for the sauce.	1.8	4.2*
The man used a basket to hold the precious liquid for the sauce.	3.0	4.0
The man used a bowl to hold the precious liquid for the sauce.	4.5	3.9
The woman used a band-aid to cover the irritated child in the waiting room.	1.5	4.2
The woman used a parody to entertain the irritated child in the waiting room.	4.0	3.2
The woman used a doll to entertain the irritated child in the waiting room.	5.0	4.7
Beatrice used a towel to dry the important program on the computer.	1.5	3.6
Beatrice used a key to open the important program on the computer.	2.0	4.8
Beatrice used a password to open the important program on the computer.	5.0	4.3
Stuart used a blender to mix the various dimensions of the poster.	1.0	4.9*
Stuart used a stopwatch to measure the various dimensions of the poster.	1.5	4.3
Stuart used a ruler to measure the various dimensions of the poster.	4.8	4.6
Nancy used a sponge to wash the long cigarette that a friend gave her.	1.3	4.1*
Nancy used a torch to light the long cigarette that a friend gave her.	3.8	4.2
Nancy used a match to light the long cigarette that a friend gave her.	5.0	4.4
The man used a formula to explain the beautiful yacht after the outing.	2.3	3.4*
The man used a ribbon to secure the beautiful yacht after the outing.	2.8	3.6
The man used a rope to secure the beautiful yacht after the outing.	4.8	3.7
Robert used a folder to file the large pheasant that weighed ten pounds.	1.0	4.4*
Robert used a hook to catch the large pheasant that weighed ten pounds.	3.3	4.9
Robert used a trap to catch the large pheasant that weighed ten pounds.	5.0	4.7
The woman used a pitchfork to carry the thick mascara in the morning.	1.0	3.1
The woman used a rag to apply the thick mascara in the morning.	2.5	3.7
The woman used a brush to apply the thick mascara in the morning.	5.0	4.6
Justin used a patch to mend the spotted greyhound that he walked.	1.0	3.8*
Justin used a joystick to control the spotted greyhound that he walked.	2.3	3.3
Justin used a muzzle to control the spotted greyhound that he walked.	4.8	3.7
Alberto used a chemical to treat the brand-new information about inventory.	1.3	3.9*
Alberto used a tunnel to access the brand-new information about inventory.	1.8	3.8
Alberto used a website to access the brand-new information about inventory.	5.0	4.2
Gloria used a mask to scare the annoying traffic on Main Street.	2.5	4.8
Gloria used a loophole to avoid the annoying traffic on Main Street.	2.8	3.5
Gloria used a shortcut to avoid the annoying traffic on Main Street.	5.0	4.7

*Note.* For each set of three sentences, the first sentence is the anomalous condition, the second is the implausible condition, and the third is the plausible (control) condition. Adjective fit refers to how well the adjective (e.g., *large* in the first set) fit with the words that preceded it.

\* Sets of sentences used in the supplementary analysis.

Received November 1, 2003  
Revision received March 17, 2004  
Accepted April 13, 2004 ■