Forward versus Backward Agreement Processing in Comprehension

Eric S. Solomon and Neal J. Pearlmutter
Northeastern University

Send correspondence to the second author:
Psychology Dept., 125 NI
Northeastern University
Boston, MA 02115

E-mail: pearlmutter@neu.edu
Phone: (617) 373-3040
Fax: (617) 373-8714

Running head: Forward vs Backward Agreement

Manuscript submitted 10/27/05; comments welcome.

Please do not cite or quote without permission.
Abstract

Two experiments examined whether the comprehension system automatically computes agreement relationships in its forward pass through a sentence, or whether instead it only computes agreement by backtracking upon encountering an overtly number-marked verb. Experiment 1 used a word-by-word ungrammaticality detection task and showed that readers were sensitive to probabilistic variation in verb number for modal verbs (e.g., would, must), which are not overtly number-marked. In sampled written and spoken corpora, some of these verbs occurred about equally often with singular and plural subjects (equibiased), while others occurred more often with singular subjects than with plural subjects (singular-biased). Experiment 2 used a word-by-word self-paced reading task and sentences containing the same verbs, and readers showed a larger subject number effect (sentences with plural subjects versus sentences with singular subjects) for singular-biased verbs than for equibiased verbs, indicating that they were computing number agreement and integrating subject number with an appropriate verb number feature even for non-overtly marked verbs, implicating a forward-tracking agreement system in comprehension.

Keywords: sentence comprehension, parsing, number agreement, subject-verb agreement, syntactic features
Perhaps the most critical property for the functioning of human language is its ability to deliver the components of a message spread across a series of elements discrete in time. This property permits substantial complexity and conveyance of sophisticated and arbitrary relationships, but it comes at a cost: Languages require their speakers to track these relationships in various ways, and formal properties of one part of an utterance can constrain the properties of other parts. One paradigm example of these kinds of relationships is long-distance dependency, in which elements separated by an arbitrary amount of intervening material are explicitly linked. In English and many other languages, one kind of long-distance dependency is number agreement between a subject noun phrase (NP) and a corresponding verb, as illustrated in (1), where the starred form of each verb is usually judged to be ungrammatical. An arbitrary amount of intervening material (e.g., *whose decisions were based on fundamentalist religious beliefs*) can appear between leader(s) and was/were, but the relationship will still hold.

(1) a. The leader was/*were responsible for the deaths of thousands of civilians.
   
b. The leaders *was/were responsible for the deaths of thousands of civilians.

Such relationships are very common in English and other languages, and language producers clearly must track them in order to generate grammatically correct utterances (e.g., Bock & Miller, 1991; Eberhard, Cutting, & Bock, 2005; Solomon & Pearlmutter, 2004; Vigliocco, Butterworth, & Garrett, 1996; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996). Whether comprehenders routinely track such information is less clear, however. Agreement information is not always available or useful, even in languages that rely heavily on it to mark relationships between constituents; and in English, which displays a relatively limited range of agreement phenomena, the information conveyed by agreement is typically redundant with phrase structure (word order) information,
when it can be computed at all (Pearlmutter, Garnsey, & Bock, 1999).

In fact, however, comprehenders do seem to track agreement in English and other languages: They are at least sometimes sensitive to real and apparent violations in a variety of experimental situations and tasks (e.g., Blackwell, Bates, & Fisher, 1996; Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996; Hagoort, Brown, & Groothusen, 1993; Kaan, 2002; Kail & Bassano, 1997; Nicol, Forster, & Veres, 1997; Osterhout & Mobley, 1995; Pearlmutter et al., 1999), displaying longer reaction times and differing brain activity to violations, even when the task only requires normal reading. The studies examining agreement comprehension have either studied overt violations (e.g., *The boy are...*) or cases of apparent but not actual violations, where material intervening between the head noun of the subject and the agreeing verb potentially creates interference in tracking number information (e.g., *The key to the cabinets was...*; Nicol et al., 1997; Pearlmutter, 2000; Pearlmutter et al., 1999). These studies have provided evidence for sensitivity to agreement, but aside from preliminary evidence supporting the idea that agreement features are processed hierarchically (Pearlmutter, 2000), they have not examined questions about the mechanisms underlying feature processing.

One very basic question is whether agreement is routinely processed by comprehenders as part of the grammatical processing system’s core procedures, or whether instead it is only computed when it is overtly available and might be useful (e.g., to resolve an ambiguity). In the former case (hereafter forward-tracking of agreement), the number of the subject NP would be computed (perhaps by hierarchical feature-passing) as that NP is encountered, during the processing system’s forward pass through the sentence. When the corresponding verb and verb phrase (VP) are encountered, the number of the subject NP would then be checked for agreement. If the subject and verb agree, processing would continue; if they do not, the system would have to do some additional work to reconcile the discrepancy.
The alternative to a forward-tracking system is a *backtracking* one. Particularly in English, where agreement is not often overtly marked and is usually redundant, the processing system might be more efficient if it only computed agreement relations when they were useful and available, or at least only when they were available; namely, cases in which both the subject NP’s number and the corresponding VP’s number could each be determined independently, from the input. During normal processing in such a system, the number of the subject NP might or might not be computed as the NP is processed, but regardless, no subject-verb checking process would occur unless the number of both the subject and the verb could be computed; and even then, number might only be checked if it would be useful to resolve an ambiguity or to support some otherwise-unclear relationship. Thus if, for example, the verb was not overtly marked (i.e., a form which could occur with either a singular or a plural subject, such as *would* or any past-tense verb in English other than the copula), the system would not attempt to check whether the subject NP and verb agreed. If the verb was marked overtly (e.g., *was/were*, most third-person present-tense verbs in English), however, the checking process would occur (after computing the subject NP’s number, if not already done), and a mismatch would result in additional processing difficulty relative to a correctly-agreeing case.

Pearlmutter et al. (1999) suggested that English might use a backtracking system as a matter of efficiency,¹ and Nicol et al. (1997) argued for such a system on similar grounds, but no empirical evidence differentiates the two. Either a forward- or backtracking system is compatible with the existing agreement comprehension data, because existing studies almost all make use of explicitly

---

¹Pearlmutter et al. (1999, pp. 447–448) also argued against what they called a “backtracking” system on the basis of their eyetracking data, but they were referring to a system which did not compute even the number of the subject NP as it was encountered, only doing so when the (inflected) verb was encountered. Pearlmutter et al.’s “backtracking” system is thus a more extreme subtype of ours and makes the same predictions for current purposes. We can only assume that they failed to appreciate our far more appropriate use of the terminology, and we use “backtracking” throughout the current work to refer to a system which checks agreement between the subject NP and verb only in cases where the verb is overtly marked.
number-marked verbs and subject NPs; in either system, agreement checking will thus occur when
the verb is encountered, and real or apparent violations will create disruptions relative to controls.
The one exception is Nicol et al.'s (1997) Experiment 3, in which participants read stimuli like
those in (2). Whole-sentence reading times did not differ for these pairs, and Nicol et al. noted that
the null effect — relative to a reliable interference effect in corresponding pairs where *was* replaced
*will be* (their Experiment 2) — suggested that agreement was not computed when the verb was not
overtly marked. However, the null effect is also compatible with forward-checking models, because
regardless of whether the subject NP is correctly determined to be singular or incorrectly computed
as plural due to interference, the agreement check will succeed for non-overtly marked verbs like
*will*.

(2) a. The author of the speech will be well rewarded.

b. The author of the speeches will be well rewarded.

Nevertheless, Nicol et al.'s (1997) Experiment 3 does provide a basis for understanding how
to differentiate forward- and backtracking systems: The most straightforward difference in the
alternatives arises when the verb is not explicitly marked, as a backtracking system in that case
will not check agreement, whereas a forward-tracking system will. While the check implemented
by a forward-tracking process in such a case should always yield a successful match, the matching
process might not always be the same.

For example, one way to think about the situation is that a non-overtly marked verb is am-
biguous, as it can take one or the other of two different number-values, analogous to semantically-
ambiguous words taking one or the other of two different meanings (e.g., *panel* in *The screen on
the panel fed the speaker his lines.* vs. *Recent appointees to the government panel were unquali-
fied.*). This possibility arises naturally out of a constraint-based lexicalist approach to sentence
processing, in which grammatically-relevant properties of words are linked to those words as parts of their lexical entries (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994). Furthermore, on such an approach, different ambiguous verbs might have different degrees of preference for singular versus plural number, as a function of prior exposure. Thus a verb like kissed, while grammatical with either a singular subject (e.g., Mitt kissed George.) or a plural one (Mitt and George kissed.), might be expected to occur relatively more often with plural subjects; whereas a verb like lied might be expected to appear more often with a singular subject (e.g., Dick lied routinely.). This information could be encoded in a verb’s lexical entry either as a pair of features (singular, plural) trading off against each other, or as a single feature (plural, if plural-markedness is correct; Bock & Eberhard, 1993; Eberhard, 1997) with a value that reflects the degree of preference for one of the alternatives. In either case, this information would come into play during comprehension on a forward-tracking account: A verb with a preference for plural number, for example, would be easier to match with a plural subject than with a singular subject, because the preferred alternative will have an advantage in activation level in the verb’s lexical entry.

On a backtracking approach, on the other hand, no matching process will occur, by definition: Checking against the subject NP’s number only occurs when the verb is explicitly marked. As a result, any empirical bias that a verb might have for one number value over another will never come into play, and processing of non-overtly marked verbs should not display any sensitivity to such a bias.

The experiments below made use of these predictions to examine whether number feature processing in English is handled with forward-tracking or backtracking. To minimize potential confounds between number-bias and meaning differences, the non-overtly marked verbs were modals (e.g., can, might, will) rather than open-class lexical cases (e.g., kissed, lied), and we identified modal
subsets which were more or less number-biased, to fit the predictions described above. The first experiment focused on determining whether or not comprehenders display sensitivity to differences in verbs’ number-marking biases at all, using a word-by-word grammaticality judgment task, and the second experiment directly examined the predictions of forward- and backtracking models in a word-by-word reading-for-comprehension task.

**Experiment 1**

This experiment compared the predictions of forward- and backtracking agreement processing models using a self-paced word-by-word grammaticality judgment task. Although this task may cause readers to focus more on syntactic properties than they would during normal reading for comprehension, it also should potentially highlight the presence of verb number-bias information if it is available in the lexicon.

To reveal the influence of the biases, verb bias was crossed with a manipulation of subject number like that in earlier studies of subject NP number interference (Nicol et al., 1997; Pearlmutter et al., 1999). All subject NPs and subject head nouns were singular, but an additional noun within the subject NP could be either singular or plural (e.g., *The key to the cabinet(s)*...). When both nouns are singular (*key, cabinet*), the subject NP number should be clearly singular; but when the intervening noun is plural (*cabinets*), the subject NP as a whole may be incorrectly determined to be plural. If readers compute and check agreement (on a forward-tracking system), this difference in the probability of construing the subject NP as singular versus plural should interact with verb number-marking bias, such that verbs with stronger preferences to be singular should show more difficulty when combined with subject NPs more likely to be considered plural.
In addition to the non-overtly marked modal conditions, we included conditions with the copula *was*. These conditions (e.g., *The key to the cabinet(s) was rusty*...) provide a baseline for the magnitude of the effect and replicate earlier work: Nicol et al. (1997) and Pearlmutter et al. (1999) both found that comprehenders had more difficulty (at the verb or the following word, or for the sentence as a whole) when the intervening noun was plural. From the perspective of either a forward- or a backtracking system, these cases should show a subject NP number effect. In a forward-tracking system, *was* is just another verb for agreement processes to check, and it happens to have an effectively 100% singular bias, so it should provide an estimate of the maximum subject NP number effect in this paradigm. In a backtracking system, *was* is overtly marked and thus, unlike the non-overtly marked modals, agreement will be checked and an apparent violation will be detected more often in the condition with a plural intervening noun; the copula conditions should be the only ones showing such an effect.

**Method**

**Participants.** Sixty-four Northeastern University undergraduates participated. In this experiment and Experiment 2, all participants were native English speakers and received either class credit or payment ($10). No participant was involved in more than one experiment.

**Modal bias computation.** To look for modals that varied in their number biases, 3 samples were collected from each of 3 different published text corpora for each of the modal verbs in Table 1. The 3 corpora used were the *Wall Street Journal* corpus (WSJ; ACL Data Collection Initiative, 1993), the Reuters corpus (Rose, Stevenson, & Whitehead, 2002), and the Corpus of Spoken Professional American English (CSPAE; Barlow, 2000). Each sample was taken by choosing a random starting
point in the corpus and examining the next 100 tokens containing the modal in question.\textsuperscript{2} The samples were chosen so that samples for a given modal from within the same corpus did not overlap. (In the case of the CSPAE, one sample was taken from each of the three sections of the corpus: the national reading and math committee meetings, the Univ. of North Carolina faculty meetings, and the Clinton administration press conferences.) For each token, the number of the subject NP for the modal was coded. Ambiguous cases in which the number could not be determined from overt marking in the sentence were not counted as part of the sample (a later token was substituted). The singular bias of each modal computed from the samples is shown separately for each corpus in Table 1. In addition, to provide a baseline for comparison, subject NP number regardless of verb was counted for the first clause in each sentence of an additional set of tokens, chosen using the same sampling procedures, except that approximately 900 tokens were sampled from each corpus.

The stability of the modals’ biases across corpora was checked by computing inter-corpus correlations with modal as the random factor ($N = 8$). The WSJ biases were reliably correlated with those in the Reuters corpus ($r = .98$, $p < .001$) and in the CSPAE ($r = .72$, $p < .05$), and the correlation between the Reuters and CSPAE corpora was marginal ($r = .69$).

From the modals in Table 1, two were selected as equibiased (can and must), and five were selected as singular-biased (could, may, might, will, and would). Should was not clearly in either category and was thus not used in constructing the stimuli. Across all corpora, the mean singular bias of the equibiased modals was 50%, while the corresponding bias for the 5 singular-biased modals was 68%. These means differed significantly ($t(5) = 6.82$, $p < .01$).

**Materials and Design.** Thirty-six stimulus sets like that shown in (3) were constructed. Each sentence began with a subject NP containing the determiner The, a singular head noun (e.g., key),

\textsuperscript{2}The only exception was must in the CSPAE, for which only 205 tokens were sampled, because not enough tokens could be found in two of the CSPAE sections.
Table 1: Modal Singular Biases (%) by Corpus

<table>
<thead>
<tr>
<th>Modal</th>
<th>WSJ</th>
<th>Reuters</th>
<th>CSPAE</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>can</td>
<td>53</td>
<td>50</td>
<td>39</td>
<td>47</td>
</tr>
<tr>
<td>could</td>
<td>69</td>
<td>72</td>
<td>57</td>
<td>66</td>
</tr>
<tr>
<td>may</td>
<td>69</td>
<td>71</td>
<td>59</td>
<td>66</td>
</tr>
<tr>
<td>might</td>
<td>77</td>
<td>76</td>
<td>50</td>
<td>68</td>
</tr>
<tr>
<td>must</td>
<td>56</td>
<td>51</td>
<td>52</td>
<td>53</td>
</tr>
<tr>
<td>should</td>
<td>61</td>
<td>59</td>
<td>51</td>
<td>57</td>
</tr>
<tr>
<td>will</td>
<td>72</td>
<td>70</td>
<td>62</td>
<td>68</td>
</tr>
<tr>
<td>would</td>
<td>77</td>
<td>75</td>
<td>71</td>
<td>74</td>
</tr>
<tr>
<td>M</td>
<td>67</td>
<td>66</td>
<td>55</td>
<td>62</td>
</tr>
</tbody>
</table>

Subject NP Baseline 63 68 74 68

Note. WSJ = Wall Street Journal. CSPAE = Corpus of Spoken Professional American English.

and a prepositional phrase (PP), which ended with either a singular local noun (e.g., cabinet; 3a, 3c, 3e) or its corresponding plural (3b, 3d, 3f). After the subject NP was a verb phrase which began either with the copula (always was; 3a, 3b); with a singular-biased modal (e.g., might; 3c, 3d) and be; or with an equibiased modal (e.g., must; 3e, 3f) and be. All versions of a given item were identical after the copula, and they all had a Yes/No comprehension question. The full set of sentence stimuli are included in Appendix A.
(3)  a. The key to the cabinet was rusty after years of disuse.
    b. The key to the cabinets was rusty after years of disuse.
    c. The key to the cabinet might be rusty after years of disuse.
    d. The key to the cabinets might be rusty after years of disuse.
    e. The key to the cabinet must be rusty after years of disuse.
    f. The key to the cabinets must be rusty after years of disuse.

In addition to the experimental items, 84 filler sentences were included, each with a comprehension question. Of the fillers, 44 were grammatical and were of similar length and complexity to the experimental items. Eight of these items began with subject NPs consisting of an NP PP sequence containing a plural head noun and a plural or singular local noun, followed by were...; the other 36 involved object/subject temporary ambiguities or comma-disambiguated controls (e.g., After the band left(,) the party went on...) for a separate experiment. The remaining 40 fillers were ungrammatical, containing violations of phrase structure (missing or swapped words) or subject-verb agreement. They were also similar in length and complexity to the experimental stimuli. Included in the ungrammatical set were 12 items with NP PP subject NPs; 4 had singular head nouns and 8 had plural heads, with balanced numbers of singular and plural local nouns. Half of these contained subject-verb agreement violations (created by using the wrong form of was/were), and half contained phrase structure violations. The other ungrammatical fillers were similar in structure to the grammatical object/subject ambiguous cases (12 items) or had a range of miscellaneous structures (16 items), and they contained either phrase structure or subject-verb agreement violations.

The experimental stimuli and fillers were combined to form 6 counterbalanced 120-item lists. Each list included all the fillers and exactly one version of each of the 36 experimental items.
Apparatus and Procedure. Participants were run individually. They read 10 initial practice items, followed by one of the lists, on an IBM-compatible computer running the MicroExperimental Laboratory (MEL) software package (Schneider, 1988). The stimuli were presented in a random order using a non-cumulative word-by-word self-paced moving window paradigm (Just, Carpenter, & Woolley, 1982) combined with grammaticality judgment. At the beginning of a trial, an item was displayed on the screen with all non-space characters replaced by dashes. When the participant pressed the space bar, the first word of the item was displayed, replacing the corresponding dashes. To read each successive word, the participant pressed either a YES or NO key (the \(<B>\) and \(<N>\) keys, respectively), which caused the current word to revert to dashes and the next word to appear in place of its corresponding dashes. The participants were instructed to press the YES key when the sentence so far made sense and to press the NO key when the sentence “stopped making sense”. They were told that they could switch from pressing NO back to pressing YES within a sentence if it began to make sense again, and the instructions gave an example sentence (viz., *This is how read to the sentences in this experiment.*) in which they might press YES until *read* appeared, then switch to NO, then switch back to YES around *the*.

After pressing YES or NO on the last word, the participant pressed the space bar. This caused the item to be replaced by its comprehension question, which the participant answered using the same YES or NO keys. The computer presented feedback about the participant’s answer to each question, and then the next trial began. Most participants completed the experiment in 35 minutes.

Results

Comprehension question accuracy and cumulative ungrammaticality judgments (the latter at each word position) were analyzed by condition for the experimental items. For each analysis, we performed two separate 3 (verb type) $\times$ 2 (local noun number) analyses of variance (ANOVAs),
Table 2: Experiment 1 Comprehension Question Accuracy (%) by Condition

<table>
<thead>
<tr>
<th>Verb Type</th>
<th>Local Noun Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singular</td>
</tr>
<tr>
<td>Copula</td>
<td>98.4 (0.61, 0.59)</td>
</tr>
<tr>
<td>Singular-biased</td>
<td>95.1 (1.09, 1.81)</td>
</tr>
<tr>
<td>Equibiased</td>
<td>96.9 (0.97, 0.83)</td>
</tr>
</tbody>
</table>

*Note. Standard errors of the mean computed from the analyses by participants and items, respectively, are shown in parentheses.*

one with participants ($F_1$) and another with items ($F_2$; Clark, 1973) as the random factor. In addition, we conducted relevant paired comparisons of local noun number within each verb type.

**Comprehension Question Accuracy**

Accuracy on the comprehension questions was very good, as shown in Table 2. Analyses showed a main effect of verb type ($F_1(2,126) = 3.71$, $MS_e = 81.19$, $p < .05$; $F_2(2,70) = 3.13$, $MS_e = 54.82$, $p < .05$), with the copula conditions more accurate than the modal conditions. There was no effect of local noun number and no interaction (all $F$s < 1).

**Cumulative Ungrammaticality Judgments**

The cumulative percentage of trials judged ungrammatical was computed at each word position in the experimental sentences, with the conditions aligned as shown in (4). At each position, the cumulative percentage was computed as the number of trials in which the participant had pressed the NO key at that word in the trial or on any earlier word (even if the YES key had been pressed for the current word), out of the total number of experimental item trials. Trials on which the
comprehension question was answered incorrectly (3.5% of all trials) and trials judged ungrammatical prior to the verb (0.9% of all trials) were excluded from the analyses. We also computed analyses on arcsine-transformed percentages (Cohen & Cohen, 1983): The participant × condition means were transformed for the $F_1$ analyses; the item × condition means were transformed for the $F_2$ analyses. These analyses revealed statistically-identical patterns to those on the untransformed data.

(4) a. The key to the cabinet(s) was rusty after years of disuse.

b. The key to the cabinet(s) might be rusty after years of disuse.

c. The key to the cabinet(s) must be rusty after years of disuse.

Figure 1 shows cumulative ungrammaticality judgment percentages by position and condition from the local noun through position 11 (some items only had 11 words). We report analyses for the copula (position 7; always was or be), which effectively examines exactly those trials judged ungrammatical at the modal or the copula and not earlier. Numerical patterns and analyses for position 6 (the modal alone) and positions 8–11 (following the copula) were largely identical to these.

At the copula (position 7), a reliable verb type × local noun number interaction was present ($F_1(2, 126) = 9.81, MS_e = 30.47, p < .001; F_2(2, 70) = 14.66, MS_e = 11.92, p < .001$). Plural local nouns yielded more ungrammaticality judgments than singulars in the copula conditions ($F_1(1, 63) = 16.81, MS_e = 64.20, p < .001; F_2(1, 35) = 24.08, MS_e = 25.15, p < .001$) and in the singular-biased conditions ($F_1(1, 63) = 5.33, MS_e = 10.16, p < .05; F_2(1, 35) = 5.63, MS_e = 5.51, p < .05$) but not in the equibiased conditions, which did not differ ($F_s < 1$). There were also main effects of verb type ($F_1(2, 126) = 6.38, MS_e = 29.17, p < .01; F_2(2, 70) = 7.40, MS_e = 14.25, p < .01$) and of local noun number ($F_1(1, 63) = 18.93, MS_e = 28.26, p < .001; F_2(1, 35) = 16.72, MS_e = 17.19,$
Figure 1: Experiment 1 cumulative ungrammaticality judgments by position and condition.

$\text{Discussion}$

The ungrammaticality judgment results revealed a pattern of sensitivity to the number-bias of the modal verbs, even though such verbs are not overtly marked for number in English. Not surprisingly, readers were most likely to believe that a sentence did not make sense when a plural local noun intervened between a singular head noun and the singular copula (was), but they were also more likely to show a local noun number effect for singular-biased modals than for equi-biased
Forward vs backward agreement

17

modals. This supports the idea that number-marking information, possibly featural, is stored with verbs in the lexicon and is accessed as part of the process of comprehending a sentence.

These results are also compatible with a forward-tracking rather than a backtracking account of number agreement processing in comprehension, because on a backtracking account, non-overtly marked verbs (e.g., modals) would never be checked against the number of the subject. On a forward-tracking account, the number of the subject NP is computed and checked against the number of the verb (or VP) regardless of whether the verb is overtly marked, as part of the usual set of operations involved in parsing. When the number of the subject and the bias of the verb are the same, the checking process will proceed more quickly than when they are different, because the less frequent number for the verb will have to be accessed in the latter case.

However, while the Experiment 1 results clearly indicate that at least some readers are at least on some occasions sensitive to probabilistic number information about non-overtly marked verbs, the strength of the claims from Experiment 1 is limited by two properties: First, the grammaticality judgment task explicitly required readers to attend to whether the sentence “made sense” at each word, and many of the filler sentences were overtly ungrammatical, either because of phrase structure or agreement violations. This may have encouraged attending to grammatical properties which might be ignored or not computed during normal reading. Second, the cumulative percentages of ungrammaticality judgments in Experiment 1 were very small (less than 10%). This is not surprising, as none of the critical sentences were in fact ungrammatical, and the manipulation of number in the subject relied on effects from the local noun — all of the subject NPs were singular and might have seemed plural only because of interference from the local noun. Nevertheless, it might suggest that verb number-bias information is only rarely considered during comprehension.

Experiment 2 addressed these two issues by changing the task to self-paced word-by-word reading, with no judgment component. In addition, the subject NPs were changed to be overtly
singular or plural, with the head and local nouns matching in number. The same sets of modals were used, so each type of modal — singular-biased or equibiased — appeared (separately) with an overtly singular and an overtly plural subject NP, and the critical plural conditions should have appeared plural on every trial, rather than just on a subset. Thus on a forward-tracking model, these trials should always require a plural value from the subject NP to be reconciled with the number of the verb. This should result in additional relative difficulty for the singular-biased modals compared to the equibiased modals. Under a backtracking model, however, as in Experiment 1, the bias of the modal should have no effect.

Experiment 2

Method

Participants. Ninety-three Northeastern University undergraduates participated, but the data from two were excluded: One participant had 78% correct comprehension performance (all other participants had over 80%); and another subject, after completing the experiment, reported having “developed a rhythm” during reading and using speed-reading techniques. The remaining 91 participants’ data were included in all analyses.

Materials and Design. A new set of 36 stimulus items including Yes/No comprehension questions was created, similar in structure to those from Experiment 1. An example is shown in (5). Each item began with The, followed by a head noun which was either singular or plural (e.g., key, keys), and then a PP containing a noun which always matched the head in number (e.g., cabinet, cabinets). The subject NP was followed by either a singular-biased (5a, 5b) or an equibiased (5c,
5d) modal, an untensed lexical verb (instead of be as in Experiment 1), and then the rest of the verb phrase. Copula conditions were not included, in order to focus on the contrast between the modal sets. The same singular-biased and equibiased modal verbs were used as in Experiment 1. The full set of stimuli is listed in Appendix B.

(5) a. The key to the cabinet might rust from the moisture near the window.
    b. The keys to the cabinets might rust from the moisture near the window.
    c. The key to the cabinet can rust from the moisture near the window.
    d. The keys to the cabinets can rust from the moisture near the window.

In addition to the experimental items, 64 filler sentences were included, each with a comprehension question. All the fillers were grammatical and were of similar length and complexity to the experimental items. Sixteen of the fillers began with an NP PP sequence like those in the experimental items but were followed by a copula-headed VP. Of the remaining 48 fillers, 36 involved object/subject temporary ambiguities as in Experiment 1, and the other 12 had a variety of structures involving sentence-initial adverbs or PPs. The experimental stimuli and fillers were combined to form 4 counterbalanced 100-item lists, each of which contained all of the fillers and exactly one version of each experimental item.

**Apparatus and Procedure.** All procedures were the same as in Experiment 1, except that the instructions to participants indicated that they should read as normally as possible. There were no references to judging whether sentences made sense, and participants used only the space bar to advance through each sentence, although they still used the YES and NO keys to answer the comprehension questions. All participants completed the experiment in less than 30 minutes.
Results

One stimulus item (see Appendix B) was dropped from all analyses, because it was incorrectly presented in an ungrammatical form. In addition, a counterbalancing error was discovered after the experiment was completed, which resulted in one of the stimulus lists containing more singular-biased than equibiliased items (24 and 12, respectively), while another list had the reverse ratio (12 and 24, respectively). The other two lists had equal numbers of singular-biased and equibiliased modals, and the ratio of singular to plural subject NPs was not affected by this problem in any of the lists. Examination of the numerical patterns separately by list revealed that to the extent that the lists differed, the two lists without the counterbalancing errors showed the clearest differences (corresponding to those reported below); we therefore included the data from all lists in the analyses.

Comprehension question accuracy and individual word reading times (the latter at each word position) were analyzed by condition for the experimental items. Separate 2 (verb type) × 2 (subject NP number) ANOVAs were conducted by participants and by items for each analysis, followed by relevant paired comparisons of subject NP number within each verb type.

Comprehension Question Accuracy

Accuracy on the comprehension questions is shown in Table 3. Analyses showed a main effect of verb type ($F_1(1, 90) = 8.18, MS_e = 77.72, p < .01; F_2(1, 34) = 4.84, MS_e = 43.58, p < .05$), with equibiliased items more accurate than singular-biased items. There was no effect of subject NP number and no interaction (all $F$s < 1).

Reading Time

Reading time means were computed for each condition at each word position, excluding trials on which the comprehension question was answered incorrectly (5.8% of all trials). To adjust for differ-
Table 3: Experiment 2 Comprehension Question Accuracy (%) by Condition

<table>
<thead>
<tr>
<th>Verb Type</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular-biased</td>
<td>92.7 (1.18, 1.37)</td>
<td>93.2 (0.92, 1.43)</td>
</tr>
<tr>
<td>Equibased</td>
<td>95.3 (0.79, 1.19)</td>
<td>95.7 (1.11, 1.12)</td>
</tr>
</tbody>
</table>

*Note.* Standard errors of the mean computed from the analyses by participants and items, respectively, are shown in parentheses.

ences in word length across conditions as well as overall differences in participants’ reading rates, a regression equation predicting reading time from word length was constructed for each participant, using all filler and experimental items (Ferreira & Clifton, 1986; see Trueswell, Tanenhaus, & Garney, 1994, for discussion). At each word position, the reading time predicted by the participant’s regression equation was subtracted from the actual measured reading time to obtain a residual reading time. Thus each participant’s mean reading time per word across the entire experiment was transformed to 0 ms residual reading time, and negative residual times indicate faster than average times. Residual reading times beyond 5 $SD$ from the corresponding condition $\times$ position cell mean were excluded, affecting less than 0.6% of the data, and all analyses were conducted on the resulting data set. Appendix C reports the raw reading times trimmed in the same manner as the residual reading times.

Figure 2 shows residual reading time by position and condition from the start of the sentence through position 9 (several items only had 10 words) and reveals two main results: First, plural subject NPs were read more slowly than singular ones throughout most of the subject NP and into the VP; and second, at the word following the modal (position 7), a plural subject NP created
additional difficulty in combination with a singular-biased modal, but not with an equibiased modal.

Analyses at each position supported these patterns: At position 1, the first word of the sentence, there were no reliable differences between conditions (all $F$s < 1.5, $p$s > .25); but in each of positions 2 (the head noun) through 5 (the local noun), plural subject NPs were read more slowly than singular subjects (all $F$s > 4.0, $p$s < .05; except for the analysis by items for position 2, which was marginal). There was no effect of verb type nor an interaction at any at these positions (all $F$s < 1.5, $p$s > .30). The same pattern appeared at the modal itself (position 6): The singular and plural subject NP conditions differed ($F_1(1, 90) = 19.06, MS_e = 3110, p < .001; F_2(1, 34) = 28.09,$)}
At position 7, the lexical verb following the modal, there was still a subject NP number effect 
\( F_1(1, 90) = 6.85, MS_e = 2436, p < .05; F_2(1, 34) = 6.18, MS_e = 1293, p < .05 \), and there 
was a verb type effect, with singular-biased modal conditions slower than equibiased ones 
\( F_1(1, 90) = 7.05, MS_e = 3161, p < .01; F_2(1, 34) = 12.83, MS_e = 885, p < .01 \). But, critically, 
an interaction also appeared at this position 
\( F_1(1, 90) = 5.71, MS_e = 2607, p < .05; F_2(1, 34) = 4.82, MS_e = 1128, p < .05 \). Paired 
comparisons showed that while the two equibiased conditions did not differ \( F_s < 1 \), the singular-biased plural condition was read more slowly than the singular-biased 
singular condition \( F_1(1, 90) = 10.81, MS_e = 2918, p < .01; F_2(1, 34) = 8.11, MS_e = 1642, p < .01 \).

Although the numerical pattern of the interaction continued into position 8, neither the interaction 
nor the main effect of verb type was reliable \( (all F_s < 2, ps > .15) \). The main effect of subject 
NP number was still present, however \( F_1(1, 90) = 4.75, MS_e = 2932, p < .05; F_2(1, 34) = 4.29, MS_e = 928, p < .05 \); and it continued (weakly) in position 9, where it was only reliable in the 
analysis by items \( F_1(1, 90) = 2.25, MS_e = 2744, p < .15; F_2(1, 34) = 5.32, MS_e = 508, p < .05 \). 
No main effect of verb type nor an interaction appeared at position 9 \( (all F_s < 1) \).

Discussion

The reading time results follow the grammaticality judgment results from Experiment 1 in showing 
that readers are sensitive to the number biases of (non-overtly marked) modal verbs, suggesting 
that this information is stored as part of each verb’s lexical entry. The experiment also showed 
that plural subject NPs are harder to process than singular subject NPs. This may simply be a 
matter of frequency, because singular subjects are more common in English than plural subjects, 
or it may be the result of additional complexity associated with the processing or representation of 
plurals (individual nouns or NPs) at any of a variety of linguistic levels.
Most critically, though, this experiment provides evidence from normal reading for comprehension, showing that readers compute agreement between subject NPs and corresponding VPs, even when those VPs are not themselves number-marked. This extends the Experiment 1 results and argues against a backtracking agreement processing system: Such a system is compatible with the main effect of subject NP number, but it predicts that agreement will only be checked when the verb is overtly marked, so it cannot explain the interaction between subject NP number and modal number bias at the word following the modal. On a forward-tracking system, however, the number of the subject NP will always be checked against or integrated with the number of the VP, and this will be relatively difficult in a case where a plural subject NP must be reconciled with a verb that has a preference for a singular subject. For the equibiased modals, on the other hand, reconciling with either a singular or a plural subject NP will be approximately equally difficult, because such verbs occur about equally often with singulars as with plurals.

One potential concern about these results is that the critical interaction does not appear until the word after the modal (i.e., position 7). However, self-paced reading tasks (among others) often cause effects to be spread downstream by a word or two. In Experiment 1, for example, while the interaction began to appear at the modal itself, it was stronger later; and Pearlmutter et al. (1999) found a mix of similar potentially slightly delayed effects of (un)grammaticality and head/local noun number mismatch. This is also the case with other syntactic effects unrelated to agreement phenomena (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997). Furthermore, Pearlmutter et al. (1999) showed that grammaticality effects based on agreement tend to be clearer on the word following the agreeing verb in eyetracking as well (but cf. Sturt, 2003). Thus the timing of the Experiment 2 pattern is within the range of other related effects. Nevertheless, the question of the timing of agreement computation is legitimate, and whether the apparent delay in the appearance of effects is created by properties of the task or instead reflects underlying processing is an open
General Discussion

Experiments 1 and 2 provide converging evidence for two novel results. First, they show that comprehenders are sensitive to probabilistic information about how likely a verb is to occur with a singular versus with a plural subject. Second, the experiments show that this probabilistic number-bias information is used during comprehension to check agreement, and that the checking process is more difficult when the number of the subject NP mismatches the probabilistic bias of the verb. Thus the size of the noun number effects (singular vs. plural local noun in Experiment 1; singular vs. plural subject NP in Experiment 2) was larger for singular-biased modals than for equibiated modals: In the equibiated cases, the verb’s (lack of) bias was equally compatible with singular and plural subject NPs, resulting in no difference in either ungrammaticality judgment or reading time; whereas in the singular-biased cases, the verb’s bias fit better with the singular noun cases than with the plural noun cases, resulting in more ungrammaticality judgments and longer reading times for the latter conditions.

The first of these results follows naturally from a constraint-based lexicalist view (e.g., MacDonald et al., 1994) which claims that grammatically-relevant information (among other types) is associated with individual words in the lexicon, and that this information specifies not only the grammatical possibilities for a word but also the relative probabilities of the different possibilities, which can then be used to select an alternative (or to rank-order alternatives) when an input involving the lexical item is ambiguous (e.g., MacDonald, 1993, for grammatical category; Argaman & Pearlmutter, 2002, Garnsey et al., 1997, for noun and verb argument structure). On this view,
verb grammatical number is just another type of information needed for processing, so the relative bias for singular versus plural is encoded as part of the lexical entry for each verb, and verbs which are not overtly marked for number are cases of ambiguous input. Ignoring for the moment the details of the mechanism which actually performs the syntactic computations, then, the current experiments can be conceptualized as investigating a kind of context × frequency interaction in lexical syntactic ambiguity resolution, with subject NP number as the relevant context variable and verb number bias as the frequency variable. MacDonald et al. (1994; see also Trueswell & Tanenhaus, 1994), following results from lexical semantic ambiguity resolution in Duffy, Morris, and Rayner (1988), suggested that when the context is available in advance, the general pattern in such cases depends on the nature of the frequency bias: When the lexical alternatives are equally common (equibiaised), whichever alternative the context supports will become preferred relatively quickly (and as quickly as for matching lexically-unambiguous cases). On the other hand, if the lexical alternatives differ in frequency (e.g., the singular-biased modals), processing will be more difficult relative to a matched unambiguous control when the context supports the less-likely lexical alternative, but as easy as a corresponding control when the context supports the more-likely lexical alternative. While the current experiments were not designed with the relevant unambiguous control conditions, the predictions derived from direct comparisons among the ambiguous conditions match well with the results.

The second main result of the experiments, that probabilistic number information is used during comprehension to check the agreement relationship between the verb and subject NP, reveals that agreement processing is forward-tracking rather than backtracking. On a backtracking system, agreement would only be checked when independent number information was available from the subject NP and from the verb (or VP). In the critical conditions in the current experiments, the verbs were not overtly marked with number information, so agreement should not have been checked
in a backtracking system, and sensitivity to the number of the subject NP should not have varied across the different modal conditions. The finding that subject NP number did influence reading time and ungrammaticality judgments for singular-biased verbs, while it did not for equi-biased verbs, indicates instead that the agreement checking process proceeds (1) even when the verb and VP have no overt number marking; (2) even in English, where agreement phenomena are relatively limited; and (3) even when the information delivered by checking agreement is redundant with respect to understanding the meaning of the sentence. In short, the agreement system is forward-tracking, and routine syntactic processing thus involves computing the number of the subject NP, computing the number of the corresponding VP, and checking whether the two correspond.

In the context of the arguments from both Nicol et al. (1997) and Pearlmutter et al. (1999) — that a backtracking system would likely be more efficient for English — our results supporting a forward-tracking system thus suggest that agreement processes are in fact an obligatory component of syntactic processing. Given that English’s relatively impoverished agreement system makes the efficiency argument for backtracking more potent than for most other languages, an even stronger (speculative) case can be made: namely, that agreement processes are an obligatory component in the comprehension of all languages. Although strong, this claim is compatible with and might be predicted from computational systems which rely on linguistic feature computation processes to compute syntax, typically through unification (Kay, 1985/1986; see Stevenson, 1994, for a related probabilistic system). Although the current experiments do not investigate the detailed nature of the computations involved in processing agreement, the very limited evidence available in comprehension (Pearlmutter, 2000) suggests that agreement may be handled by passing features along links in a hierarchical structure, as would be the case in feature unification systems. And stronger evidence exists for such a system in language production (e.g., Franck, Vigliocco, & Nicol, 2000).

The apparent obligatoriness of agreement computation, however, raises a separate question,
about its timing. In both experiments, the relevant bias effects were clearest on the word following
the modal rather than on the modal itself. One possibility is that this is a reflection of the behavioral
tasks used in these experiments and in others: Even if grammatical computations are performed
within 150–300 milliseconds of a word’s appearance — the earliest range in which sensitivity to
syntax has been claimed in ERP measures, for example (Friederici, Pfeifer, & Hahne, 1993) — the
downstream effects of these computations on behavior (eye movements, button presses, etc.) might
take some time to accumulate to a reliably measurable level. This situation would obviously be
compatible with feature unification or feature processing in general as an obligatory component
of syntactic computation. However, even if feature processing turned out to occur somewhat
later than the earliest syntactic processes (e.g., construction of phrase structure), this would not
necessarily be a problem for the idea that feature processing is an obligatory component of syntactic
processes. In fact, to the extent that processes like feature unification are structure-constraining
rather than structure-generating, they require one or more candidate phrase structures before they
can operate. For subject-verb agreement in particular, syntactic paths for feature-passing must be
in place, linking the noun head of the subject NP with the subject NP node itself, and linking the
number-marked verb with the VP node. In a forward-tracking system, the noun links to the NP
will be created and the NP’s number features will be computed as the subject NP is processed; but
the link involving the verb will only be completed when the verb is encountered in the input. Once
the verb’s lexical entry is accessed, including its grammatical category and number feature(s), it
will be connected into the existing phrase structure, and at this point feature-passing can be used
to check subject-verb agreement. Thus even if feature-based processes like subject-verb agreement
are obligatory syntactic operations, they may be contingent on other earlier ones, such as phrase
structure building, and thus they may be delayed relative to a particular word’s onset and may not
have progressed sufficiently during processing of that word to influence behavioral responses to it.
reliably.

While the current results are clearly compatible with feature unification as the mechanism for computing subject-verb agreement, the finding that number-ambiguous verbs are lexically encoded with information about their alternative number possibilities has an interesting additional theoretical consequence for feature unification theories, and specifically for those which use feature unification as part of a constraint-satisfaction process verifying the legitimacy of a structural representation for a given input. In particular, some theories of this sort (e.g., Pollard & Sag, 1994) rely on underspecification to simplify lexical representations. In such theories, the lexical entry for a modal verb like *can* makes no reference to number information, because *can* does not constrain the number of its subject; only verbs with restrictions contain any reference to the restricting information (e.g., the lexical entry for *kisses* is explicitly linked to the combination of singular number and third person). From the perspective of constraint satisfaction, underspecification simplifies lexical entries: Any property not listed as a requirement (i.e., a constraint) is implicitly permitted, and thus in the representation of a particular sentence, *can* (for example) will end up with whichever number value the subject requires. The problem is that if non-restricting verbs contain no reference to possible number values, the relative probabilities of the possibilities cannot be tracked. Thus the current results suggest that underspecification is not an option for lexical representations, at least for information for which the processing system is sensitive to relative probabilities. This limitation on underspecification would presumably also hold for other cases in which lexical information is the source of probabilistic biases but a particular lexical form is compatible with all of the possible values for a particular feature. This may or may not encompass a wide range of information beyond grammatical number, but MacDonald et al. (1994) suggest that lexical frequencies are tracked for every type of lexical information needed for computation of syntax (cf. Gibson, in press, and Jurafsky, 1996, for effectively comparable approaches); if this turns out to be correct, underspecification
would not be available to syntactic theories intended to reflect lexical knowledge, at least as it is used in processing.

Regardless of their implications for particular syntactic theories, the current results, coupled with those from prior work, suggest that at least in English agreement processing, number information is passed hierarchically along the phrase structure being built for an input, and that feature-passing takes place as the phrase structure and feature information become available (i.e., the process is forward-tracking). The number feature information comes from lexical knowledge of individual heads and is passed up or down the relevant structure, with comparisons taking place where specified by the syntax; in the case of subject-verb agreement, the comparisons might always occur as number information is passed up from both the subject NP and the corresponding VP to the immediately dominating clausal node (S, IP, etc.), or the comparisons might occur at whatever point the subject number and verb number information happen to meet during the course of processing. In the latter case, for example, because the subject NP will be encountered earlier, its number information may be carried up through the subject NP node and then down to the anticipated VP node or even an anticipated Verb (V) node, so that the comparison would take place at the V or VP node, once the number information for the input verb was accessed from the lexicon. No existing data reveal which of these possibilities is correct; these details of the feature-passing process remain to be investigated. The current results also show that the number information used in this process is not simply one or more present or absent features, but is instead probabilistic information about the possible features; and the comparison process, wherever it occurs, is a matter of reconciling different probabilities rather than just unifying (or failing to unify) features — the process involves combination of values rather than just comparison (cf. Gibson, in press; Jurafsky, 1996; Stevenson, 1994).

A final question, so far left completely open, is why some of the modals we examined were,
in fact, equibiased, while most were singular-biased. We had no a priori basis for expecting any modals to be more or less singular-biased than others, and the source of the differences in bias remains unclear. The biases are not entirely random, as they were correlated across corpora; in other words, most modals’ relative biases appear to be fairly stable. However, it remains to be determined whether there is an underlying explanation for why some modals (viz., can, must) are less strongly biased than others, or whether instead the source of the variation is just small random differences in distribution reinforced over time. For some kinds of lexical frequency biases, the source can be traced to language-external properties such as the nature of physical reality, cultural patterns, or human social behavior (e.g., Argaman & Pearlmutter, 2002). Thus, for example, ball is presently more likely to refer to an object used in a sport or game than to a formal dance, because the former is more common in current society. Similarly, kissed is probably more likely to take a plural subject than lied because of the physical and cultural properties of the respective events they refer to. However, unlike open-class verbs such as kissed and lied, we are not aware of any language-external connection between subject numerosity and the meaning properties specified by modal verbs (e.g., possibility, necessity, obligation, permission, volition; Huddleston & Pullum, 2002; Quirk, Greenbaum, Leech, & Svartvik, 1985). In fact, can and must do not clearly group together and simultaneously separate from other modals on any of the syntactic, semantic, or pragmatic properties typically identified with the modal auxiliary lexical category; though two potential connections to consider are their use in command indirect speech acts (e.g., You can/must go outside now.) and their interactions with negation (viz., It must be the case that not-X. can be paraphrased It cannot be the case that X.) (e.g., Quirk et al., pp. 147, 219–236). A more likely possibility is that the modals vary along a continuum in their singular bias as a consequence of a variety of factors rather than any single one; this remains a topic for future investigation.

The current results thus provide evidence about one particular type of long-distance dependency
and how it is tracked in English. Subject-verb agreement, like other examples of such dependencies, requires the comprehension system to reconcile information from separate parts of an utterance, potentially spread out over time. Although English displays relatively limited overt agreement phenomena, it makes extensive use of long-distance dependencies in general (e.g., wh-questions, relative clauses) and of languages’ ability to spread information across an utterance in a range of ways; indeed, the limited use of overt agreement makes the apparent obligatoriness of its tracking all the more striking. To the extent that different long-distance phenomena are handled through the same types of mechanisms and interactions between working memory, the lexicon, and knowledge of syntax, we should expect to observe similar results in other languages and for other types of dependencies.
References


Barlow, M. (2000). *Corpus of Spoken Professional American English* [CD-ROM]. Houston, TX: Michael Barlow [Producer], Athelstan (www.athel.com) [Distributor].


Appendix A: Experiment 1 Stimuli

The singular local noun versions of the Experiment 1 stimuli are shown below with their singular-biased and equibiased modals (preceding and following the slash, respectively). The versions with \textit{was} were created by dropping the modal and changing \textit{be} to \textit{was}.

1. The slogan on the poster might/must be chanted during the demonstration.

2. The key to the cabinet might/must be rusty after years of disuse.

3. The door to the office will/must be locked to ensure the utmost privacy.

4. The check from the stockbroker will/can be cashed at the bank across town.

5. The problem in the school would/can be discussed at the meeting.

6. The defect in the car might/can be fixed by the dealer at the end of the month.

7. The mistake in the program might/must be due to careless editing by the printshop.

8. The crime in the city could/must be at an all time low.

9. The memo from the accountant may/must be sent to everyone in the department.

10. The letter from the lawyer would/can be delivered earlier than promised.

11. The warning from the expert could/must be taken seriously by the security guards.

12. The picture on the postcard would/must be cherished by the tourists.

13. The label on the bottle could/can be changed by the head of the company.
14. The name on the billboard will/must be recognizable to nearly everyone.

15. The bridge to the island will/must be soaked during the summer rains.

16. The entrance to the laboratory might/can be blocked from the inside.

17. The photo in the magazine could/must be consulted every day in the lawsuit.

18. The assistant to the manager could/can be accused of embezzling funds.

19. The pond near the trail may/must be visible even in the old photograph.

20. The design on the kite will/can be noted for its unique use of color.

21. The memo to the committee may/can be passed on to all of the workers.

22. The rash on the baby might/can be expected to disappear before the weekend.

23. The prize for the winner may/must be different from the prize in earlier years.

24. The article about the movie would/can be criticized for its use of sexist
   
25. The videotape of the play will/can be described as careless and sloppy.

26. The cushion for the cat may/can be hidden behind the heavy curtain.

27. The painting of the house will/must be cleaned carefully before the estate sale.

28. The entree on the menu could/can be advertised as low-calorie and delicious.

29. The review of the book might/must be incredibly long and unbearably complex.

30. The cord for the stereo could/must be called back due to a technical oversight.

31. The bone for the dog may/can be found out on the neighbor’s lawn.
32. The patent for the invention would/must be expiring within the next two years.

33. The game in the arcade may/can be fun for children of all ages.

34. The logo on the envelope would/can be printed with a special five-color process.

35. The creek below the bridge might/can be nearly dry before the storm season.

36. The zipcode on the package would/must be read out loud by the postal worker.
Appendix B: Experiment 2 Stimuli

The singular versions of the Experiment 2 stimuli are shown below with their singular-biased and equibiased modals (preceding and following the slash, respectively). The plural versions were created by changing both the head and local noun into their plural forms. Other differences between versions of an item are noted where relevant. Item #10 was excluded from all analyses because it was ungrammatical as presented (wear incorrectly appeared twice in a row).

1. The key to the cabinet might/can rust from the moisture near the window.
2. The door to the office could/must creak because of misaligned hinges.
3. The picture on the postcard would/can fade from laying in direct sunlight.
4. The kitten on the cushion may/must purr loudly when stroked.
5. The pond near the trail could/can flood rapidly if there is an intense storm.
6. The DJ on the turntable will/must spin some records for the large crowd.
7. The assistant to the executive will/can laugh at the reporters’ foolish questions.
8. The model on the poster might/must smile brightly while posing for the camera.
9. The secretaries to the supervisors would/can yell at the customers if they were/are unruly.

(Were was used with would and are with can.)
10. The tourist with the suitcase could/must wear brightly colored socks to stand out from the rest of the crowd in the airport.
11. The soldier with the wound might/can exercise more often to build up his/their strength.

   (His was used with singular subject NPs and their with plurals.)

12. The actress near the car would/must answer the detective’s questions when necessary.

13. The parrot on the newspaper may/can imitate the annoying housewife for hours.

14. The martian by the boulder would/must wave goodbye if the spaceship departed.

15. The psychic in the commercial could/can claim to be the least expensive.

16. The athlete by the bench could/must attempt to lift the heavy weights.

17. The singer for the performance will/can lose the competition if she/they scream(s) on the roller coaster. (She screams was used with singular subject NPs and they scream with plurals.)

18. The surgeon for the transplant may/must request a third nurse in the procedure.

19. The teacher of the class may/can argue with the parents of the children who do poorly.

20. The character in the story would/must fight the dangerous villains near the very end.

21. The musician with the earring will/can smoke cigarettes while playing.

22. The painter near the swing will/must rinse the brushes in turpentine before heading home.

23. The check from the stockbroker might/can sit on the counter for another two weeks.

24. The mistake in the program could/must cause the printing company a lot of embarrassment.

25. The slogan on the poster could/can infuriate the mayor if done in poor taste.

26. The videotape of the show may/must play during happy hour at the bar.
27. The courier with the document will/can travel through Germany and France next week.

28. The student from the newspaper might/must interview the head of the department.

29. The author of the article may/can complain about the state of the modern novel.

30. The caterer for the party will/must offer the guests a wide variety of foods.

31. The monkey in the cartoon might/can steal the banana from the unsuspecting caretaker.

32. The gardener beyond the hedge would/must search for the clippers if they were lost.

33. The surfer with the trophy may/can keep out of the water if the waves start to look too dangerous.

34. The florist with the discount will/must arrange the flowers at the funeral home.

35. The spider under the box would/can weave a web to catch the circling flies.

36. The counselor with the tattoo might/must introduce the guest speaker to the eager crowd.
Appendix C: Experiment 2 Trimmed Raw Reading Time

Table C.1: Experiment 2 Trimmed Raw Reading Time (ms) by Position and Condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular-Biased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singular</td>
<td>342</td>
<td>382</td>
<td>371</td>
<td>329</td>
<td>370</td>
<td>392</td>
<td>386</td>
<td>373</td>
<td>361</td>
</tr>
<tr>
<td>Plural</td>
<td>346</td>
<td>405</td>
<td>393</td>
<td>337</td>
<td>407</td>
<td>421</td>
<td>414</td>
<td>388</td>
<td>376</td>
</tr>
<tr>
<td>Equibased</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singular</td>
<td>340</td>
<td>376</td>
<td>358</td>
<td>323</td>
<td>364</td>
<td>378</td>
<td>372</td>
<td>367</td>
<td>359</td>
</tr>
<tr>
<td>Plural</td>
<td>336</td>
<td>396</td>
<td>379</td>
<td>332</td>
<td>382</td>
<td>399</td>
<td>372</td>
<td>373</td>
<td>364</td>
</tr>
</tbody>
</table>

Note. Reading times were trimmed at 5 SD. Position 6 is the modal verb.
Author Note

This research was supported by NSF grant BCS-9729007 and NIH grant R01 DC05237-01 to NJP. We thank Michael Amato, Vered Argaman, Mathew Arruda, Aimee Bittinger, Joanna Currier, Grazyna Furman, Yuka Kominato, Joanne Malinka, Kimberly Mitchell, Joey Pasquino, Shelagh Scollin, Elizabeth Tatro, and Naa-Solo Tettey for aid in conducting the experiments, and Michael Amato in particular for help in preparing the manuscript. Portions of this work were presented at the 2003 CUNY Sentence Processing Conference (Cambridge, MA) and the 2003 Psychonomic Society Conference (Vancouver, BC).