

## Subject–Verb Agreement Processes in Comprehension

J. L. NICOL, K. I. FORSTER, AND C. VERES

*University of Arizona, Tucson*

Studies of elicited sentence production show that the occasional subject–verb agreement errors that speakers make are more likely to occur when a singular head noun is followed by a plural, as in *The producer of the adventure movies have arrived*, than when a plural head is followed by a singular (e.g., Bock & Miller, 1991). The significance of this asymmetric pattern of errors depends on whether interference from plurals arises only during the production of sentences, or whether it also occurs in sentence comprehension tasks. Five reading experiments revealed the following: (1) patterns of reading times mirror the production error asymmetry; (2) a phrase which is *conceptually plural* but grammatically singular (e.g., *The label on the bottles*) produces no more reading difficulty than one which is conceptually and grammatically singular, a result which mimics Bock and Miller's 1991 production results; (3) interference from an intervening plural depends on a close syntactic link to the head noun phrase (e.g., The owner of the house who charmed the realtors). These results suggest that although the computation of agreement may be accomplished differently in the two systems, interference may arise whenever a structure containing a singular head and intervening plural is computed, whether during production or comprehension. © 1997 Academic Press

Most human languages exhibit agreement of one sort or another; for example, many languages require that particular elements in a sentence agree in terms of a specific feature, such as number, gender, and/or animacy. Typically, one element *controls* agreement on a later-occurring one. Because the agreeing ele-

ments are often separated by intervening words and phrases which may themselves bear agreement features, agreement phenomena pose a challenge to the sentence processing systems. From the standpoint of sentence production, the challenge is to produce the correct form of the *dependent* or *controlled* element. With respect to comprehension, the possibility of intervening elements means that the processor may need to figure out which items must agree and which do not: This would require computing, in a timely fashion, the syntactic relations among the phrases.

This research was conducted while the first and second author were supported by Grant DC-01409, a Research and Training Grant funded by the National Institute on Deafness and Other Communicative Disorders to the National Center for Neurogenic Communication Disorders, University of Arizona. During this time, the third author was supported by the Cognitive Science Program, University of Arizona. We thank Brian Butterworth and two anonymous reviewers for comments on a previous version of this article. We are grateful to Jason Barker for testing subjects, preparing stimuli and analyzing data, and Martha Barron, Nicole Diamond, Lea Ann Hald, and Tracy Love for testing subjects. Thanks also to Neal Pearlmuter for discussing with us the results of his experiments with Kay Bock and Susan Garnsey. The results of Experiment 1 were presented at the Seventh Annual CUNY Conference on Human Sentence Processing, New York, NY, March, 1994.

Address reprint requests to Janet Nicol, Psychology 312, University of Arizona, Tucson, AZ 85721. E-mail: nicol@aruba.ccit.arizona.edu.

Consider, for example, subject–verb agreement in English and a number of other languages. While the verb may be contiguous with the subject, as in sentences such as (1), sentences (2) and (3) show that it needn't be. In principle, there is no limit on the length of the intervening material.

- (1) The students are protesting.
- (2) The students from the university are protesting.
- (3) The students from the university in the outskirts of the city by the lake are protesting.

In this paper, we examine the comprehen-

sion of sentences containing a subject noun phrase (NP) containing several daughter NPs, such as *The students from the university*, as in (2). Past research on this topic has concentrated on production processes. That work has revealed a striking finding: an asymmetry in the processing of singular vs. plural forms. Bock and her colleagues (Bock & Miller, 1991; Bock & Cutting, 1992; Bock & Eberhard, 1993), and others (Vigliocco, Butterworth, & Semenza, 1995; Vigliocco, Butterworth, & Garrett, 1995; Vigliocco & Nicol, 1997) have examined the production of subject–verb agreement errors by means of a task in which participants in the experiment hear or read a sentence fragment, repeat the fragment, and complete the sentence. Sentence fragments typically contain a singular or plural “head” noun and a singular or plural “nonhead” noun; for example, *The key to the cabinet*. . . . Significantly more errors occur when the two nouns mismatch in number. However, there is an asymmetry in the two mismatch conditions: The condition which induces by far the greatest number of errors is the condition in which the head noun is singular and the nonhead plural (as in, *The key to the cabinets are shiny*).<sup>1</sup> The explanation offered by Bock and Eberhard (1993) and Eberhard (1993) for this error asymmetry has to do with the markedness of the plural form. They claim that the singular is unmarked with respect to number, but the plural is marked. Normally, plural heads transmit their plural feature to the verb, but singular heads do not; hence, the verb is either marked as plural or remains unmarked, and if the verb is unmarked, it is interpreted as singular. An agreement error occurs when the number feature of the nonhead is erroneously transmitted to the verb. Since only plurals transmit features, errors occur only when the nonhead is plural.

It has also been found that, in English, con-

ceptual factors do not play a role in the production of subject–verb agreement. Bock and Miller (1991), for example, compared relative incidence of subject–verb agreement errors following *single-token* preambles such as those in (4) with *multiple-token* preambles such as those in (5).

- (4a) SS The key to the cabinet . . .  
 (4b) SP The key to the cabinets . . .  
 (5a) SS The label on the bottle . . .  
 (5b) SP The label on the bottles . . .

In a sentence such as (5b), *the label* plausibly represents multiple tokens (i.e., one label per bottle, hence multiple labels); the head noun is grammatically singular, but *notionally* plural. In (4b), however, *one key* could be associated with multiple cabinets; *the key* refers to a single token. Bock and Miller found that multiple-token preambles did not increase the incidence of agreement errors. Sentences such as (5b) did induce errors, but at no higher rate than did the single-token preambles. This finding of an equivalent mismatch effect suggests that the computation of subject–verb agreement makes reference only to *grammatical* number, not to *notional* number. (Interestingly, this effect does not hold for languages like Italian; Vigliocco et al. (1995) found that multiple token items *do* elicit a greater number of agreement errors than single token items.)

It is not known precisely how the plural feature on the nonhead is transmitted to the verb. There is suggestive evidence, however, that feature transmission is not rightward, directly from the nonhead, or “local” noun, to the verb, but rather leftward, from the local noun phrase to the head noun phrase, and then to the verb. In the relevant studies (Vigliocco & Nicol, 1997; Nicol & Vigliocco, 1997), syntactic distance from the mismatching noun (underlined in the examples below) to the head noun was manipulated as shown in (6) and (7):

- (6a) The telegram to the friends of the soldier . . .  
 (6b) The telegram to the friend of the soldiers . . .

<sup>1</sup> The strength of the asymmetry may differ cross-linguistically. Although it appears to be quite robust in English (e.g., Bock & Miller, 1991) and Spanish (Anton-Mendez, 1996), it seems to be virtually nonexistent in Italian (Vigliocco, Butterworth, & Semenza, 1995).

(7a) The experiment which the students ran . . .

(7b) The experiment which bored the students . . .

These studies found that the more deeply embedded the mismatching NP (as in the (b) versions), the fewer the errors. This was so despite the fact that the most deeply embedded mismatching NP was, in the sentences produced by the speakers in those experiments, contiguous with the verb. One interpretation of this is that the plural feature erroneously percolates up the syntactic tree from the non-head NP to the head NP.

There appear to be limits on how far from the head the mismatching NP can be in order for interference to occur. The number of errors following sentences such as (6b) was especially small, suggesting that the number of the verb was already specified by the time the final NP was uttered. In other words, sometime after the speaker utters the head noun, the verb number is specified, even if the complex subject NP is still in the process of being output. This finding was replicated by Nicol (1995), who examined sentences such as the following, in which the attachment of the modifying clause was manipulated:

(8a) The owner of the house who charmed the realtor . . .

(8b) The owner of the house who charmed the realtors . . .

(9a) The owner of the house which charmed the realtor . . .

(9b) The owner of the house which charmed the realtors . . .

In (8), the relative clause modifies the head; by contrast, in (9), the relative clause modifies the NP within the modifying PP. Note that, in both of the (b) examples, there is an equivalent *linear* distance between the mismatch and the head NP, and between the mismatch and the verb. Error rates were surprisingly low, and did not differ across the two sentence types. The mismatch conditions together yielded an error rate of under 2%, far less, for instance, than the 12% found by Bock and Cutting

(1992) for simpler subject NPs like *The editor who rejected the books*. The very low error rate and the lack of an attachment effect are consistent with the notion that there is a limited processing window during which the head NP is active, and that verb number is specified during this time. Later-occurring mismatching NPs may not be active during this critical window.

In all, the production findings suggest that subject-verb agreement in English is a syntactically based operation which works in one of two ways, depending on the head noun number. (1) If the head noun is singular, it is unmarked. No feature is transmitted to the verb. An unmarked verb is assumed to be singular, and the singular form is produced. (2) If the head noun is plural, it is marked. The plural feature is transmitted from the head noun phrase to the verb, and the plural form of the verb is produced.

Errors such as *The key to the cabinets are here* would be produced as follows. The plural local noun (*cabinets*), being marked, bears the plural feature. The plural feature is erroneously transmitted to the head noun phrase. The head noun phrase transmits the plural feature to the verb, and the plural form of the verb is produced. The farther the mismatch from the head NP, the smaller the chance of interference.

As described, it would seem that the process of feature transmission—and erroneous feature transmission—would be unique to the production system. But it is possible that the potential for feature migration arises *whenever* a syntactic structure is computed; whether for the purposes of sentence production *or* comprehension. In other words, the computation of syntactic structure may routinely involve the identification of agreement features and these features may be “slippery,” so that they become associated with the wrong syntactic constituent. If this possibility is correct, then it ought to be the case that feature migration effects can be observed in comprehension tasks as well as production tasks. This issue is the focus of the present paper.

There are a number of reasons to think that

feature migration is purely a production phenomenon. One reason is that production and comprehension systems are logically different: They have different goals and require different processing routines. Such differences are manifest in aphasia; for example, production may be impaired in the face of preserved comprehension, or vice versa (e.g., Goodglass & Kaplan, 1972). Child language may also show comprehension/production asymmetries. For example, Keeney and Wolfe (1972) showed that children who correctly produced subject–verb agreement failed on comprehension tasks which called upon sensitivity to subject–verb agreement.

Another reason to think that feature migration is a production phenomenon is that comprehenders may not always attend to subject–verb agreement marking. Frazier (1987), for example, presented Dutch-speaking subjects with relative clause sentences which were temporarily ambiguous between a subject relative clause and object relative clause and which were disambiguated by marking on the sentence final verb. Their answers to comprehension questions about the object relatives indicated that they ignored subject–verb agreement approximately one-third of the time. This could indicate that subject–verb agreement may not always be computed during on-line sentence comprehension, even in Dutch, a language which marks verb number to a much greater extent than English.<sup>2</sup>

Certainly, it would not be too surprising to find that in certain languages, subject–verb agreement is not routinely computed during sentence comprehension. In English, for instance, not only is overt subject–verb agreement relatively infrequent (i.e., it occurs only with the third person, and only in the present tense and in the past tense with the copula), but it does not provide information about

verb–argument relations the way it does in other languages. Typically, it is clear from the structure of a clause which is the head NP of the subject: the head NP is, roughly, the highest NP within the subject NP. But just because subject–verb agreement is not usually necessary in order to compute predicate–argument relations, this does not mean that perceivers do not pay attention to such information, or that they do not to use it, for example, to confirm their syntactic assignments. Further, there are cases in which perceivers must pay attention to verb number, since it provides critical disambiguating information as to structure. For example, verb number dictates the correct structure of ambiguous possessives like *Mary and John's mother is/are . . .*, and attachment ambiguities in relative clause constructions (such as *the woman with the children who is/are . . .*). And in fact, a number of experiments have shown that subject–verb agreement does *not* go unnoticed by comprehenders. For example, Freedman and Forster (1985) found that sentence matching times were slower for sentences containing subject–verb agreement errors. Further, Osterhout and Mobley (1995) have reported that such errors elicit a *P600*, the event-related potential (ERP) component that has been shown to be associated with syntactic anomaly (Friederici & Mecklinger, 1996; Hagoort, Brown, & Groothausen, 1993; Neville et al., 1991; Osterhout & Holcomb, 1992). Subject–verb disagreement effects for constructions containing a complex subject NP have also been observed with a variety of other techniques. For example, Pearlmutter, Garnsey, and Bock (1995) have found a disagreement effect as measured by reading latencies, with both eye-tracking and self-paced reading techniques; Sevald and Garnsey (1995) combined self-paced reading of the complex subject NP with naming, and found slower naming times to the verb when it disagreed in number with the head NP; Gorrell (personal communication) found longer lexical decision times to a disagreeing verb in a syntactic priming experiment.

But even if comprehenders do compute subject–verb agreement systematically, it is rea-

<sup>2</sup> Alternatively, Frazier's results could suggest that subject–verb agreement is ignored only in the face of a strong structural bias; readers may notice the number marking, but rather than forcing a reanalysis of the sentence, the verb might be assumed to simply have the wrong inflection.

sonable to suppose that the mechanisms underlying agreement are different in comprehension and production. We assume that the production mechanism computes the structure of a subject NP and specifies the form of an upcoming verb through feature transmission. It is unlikely that a comprehension device would operate this way, since hypothesizing the number of an upcoming verb has little point if the input does not contain an overtly number-marked verb. Rather, readers and listeners may compute the structure of a subject NP, but check the head noun number *only* with the appearance of an inflected verb. Since, on this scenario, there is no number transmission to a verbal or inflectional node, there may be no inadvertent number percolation. In sum, if the mismatch effect is entirely a consequence of the way in which verb number is specified during sentence production, then this effect may not appear in comprehension.

Another scenario is also feasible. As we described above, it is also possible that *when-ever* a syntactic structure is computed, the migration of features may occur. On this view, any short-term memory representation of a syntactic structure in which some of the constituents are marked with agreement features is potentially vulnerable to feature slippage. Presumably, some type of sentence representation is constructed during the comprehension of a sentence; it is reasonable to assume that syntactic aspects are part of such a representation. If the mismatch effect is due to the percolation of features within a structured memory representation, then we should observe a mismatch effect in comprehension. Further, we should find that lengthening the distance from the mismatch to the head NP should reduce the mismatch effect, as it appears to do in production.

The purpose of the present study is to test whether a mismatch effect can be found in a comprehension task. If a mismatch creates uncertainty about NP number for the perceiver (here, the reader), and if such uncertainty leads to processing difficulty, then we expect readers to show difficulty in just the condition in which speakers make agreement errors,

namely, the singular head, plural local noun condition. Experiments 1 and 2 use different techniques to test reading times for sentences containing complex NPs such as the sentences in (4) above. In Experiment 3, inflected verbs are replaced with uninflected forms to verify that the effects seen in Experiments 1 and 2 are indeed due to agreement. Experiment 4 explores whether notional plurality of the subject NP (as in example (5)) produces an additional effect. Finally, Experiment 5 explores whether manipulating the syntactic distance between the head NP and a plural nonhead NP, as in examples (8b) and (9b), has any effect on comprehension.

The first experiment used a self-paced reading task that requires a word-by-word commitment to a particular interpretation. The standard word-by-word reading task may encourage, but does not force, the reader to make such a commitment, since a "wait-and-see" strategy can be employed. The technique, which we call the Maze task, was first used by Freedman and Forster (1985) to study the processing of sentences containing a subadjacency violation. In this task, the participant is presented with the first word of the sentence, and is then given two alternative continuations of the sentence, only one of which is grammatical. The participant must decide which of these two words is the better continuation for the sentence. This decision is indicated by pressing one of two response keys. As soon as the decision is made, another pair of alternatives is presented. In some cases, the better continuation will be the right-hand word, in others, it will be the left-hand word, and thus the participant "winds" her way through the sentence maze, making one choice at a time, with each pair of alternatives disappearing once a choice is made. If the wrong word is chosen, the sentence stops mid-stream, a feedback message signaling an error appears, and the participant advances to the next sentence. An example sentence appears in Fig. 1.

It should be noted that the only RT of interest is that of the response to the verb (i.e., *was* in the example), since it is at this point that agreement first affects decision making. It was

The	...
hesitates	announcement
pillow	by
the	door
instead	director
was	ink
disturbing	elephant
to	egg
under	everyone

FIG. 1. Illustration of the *Maze Technique*. The participant saw successive pairs of words, one pair at a time, starting with "The . . .". Thereafter, the participant pressed a right-hand or left-hand response key to indicate which word represented the better continuation. In this example, the sentence is: *The announcement by the director was disturbing to everyone*. Reading times for the pair containing the verb *was* were recorded.

also always the case that the verb agreed with the subject, and most importantly, the alternative to the verb was typically from a different lexical category. Thus, the choice was not between a verb that agreed in number with the subject and a verb that did not. It was a choice between a verb (which happened to be inflected for number) and a non-verb. Thus, the task avoids focusing attention on agreement *per se*.

The emphasis on a word-by-word commitment in this task can be seen by considering the situation faced by a participant who has correctly chosen *announcement* as the first choice. There is no way that the subject can discriminate between *pillow* and *by* unless it is seen that *announcement* is part of an NP which can be continued by a prepositional phrase but not by another noun. Similarly, there is no way to decide between *the* and *door* unless it is recognized that *by* introduces a prepositional phrase. Hence, participants cannot postpone fully comprehending the sen-

tence, because they must make word-by-word decisions about the structure of the sentence. For this reason, we can pinpoint exactly where we expect to see an effect of processing difficulty, namely, at the verb. Since this study was exploratory, we needed to be sure that we would detect an effect if an effect was to be detected.

## EXPERIMENT 1

The question addressed in this preliminary experiment was whether the conditions that give rise to subject-verb agreement errors in a sentence production task would also produce an increase in processing time in a sentence comprehension task.

### Method

*Participants.* Thirty-three University of Arizona undergraduates participated in this experiment for course credit.

*Materials.* Forty-eight quadruplets were constructed, in which the head and nonhead were either singular or plural. All sentences contained nine words. Examples are shown in (10), along with the sentence codes which will be used throughout this paper (SS = singular head, singular local noun; SP = singular head, plural local noun; PP = plural head, plural local noun; PS = plural head, singular local noun).

(10a) SS The author of the speech is here now.

(10b) SP The author of the speeches is here now.

(10c) PP The authors of the speeches are here now.

(10d) PS The authors of the speech are here now.

The experimental sentences were counter-balanced across four presentation lists so each sentence variant appeared in a different list and equal numbers of tokens of each sentence type (twelve tokens per type) appeared in each list. In addition, there were 18 distracter sentences which varied in content and syntactic structure. In all, there were 66 trials, along with 10 practice items.

*Procedure.* As described above, sentences were presented visually in the following fashion: the first word (e.g., *The*) appeared to the left of center and a series of four dots appeared to the right. The participant pressed a left-hand response key to advance to the next word. All subsequent words were paired with unrelated words which differed in grammatical category and could not continue the sentence grammatically. Pairs of words appeared on a computer screen for 500 ms and then disappeared. Participants were required to read each word pair, determine which of the two words represented a better continuation of the sentence, and press one of two response keys, whichever corresponded to the better continuation (the left-hand response key for word on the left, and the right-hand key for the word on the right). The next pair of words appeared only once participants had pressed a response key. If the participant made an error, the sentence stopped abruptly, an error message was displayed, and the participant then advanced to the next item. Each new sentence was initiated by the press of a footpedal.

The items were presented on a computer-controlled video display using the DMASTR system developed by K. I. Forster and J. C. Forster at the University of Arizona. This system synchronizes the display code with the video raster, allowing accurate reaction time (RT) measurement.

### *Results*

For this and all subsequent experiments, data from trials in which an incorrect response was recorded were discarded. RTs were trimmed, such that RTs greater than two standard deviations from the subject's mean for that condition were replaced with the appropriate cutoff value. Subjects who made more than 20% errors were replaced. In this experiment, only one subject's data were excluded due to error rate, and 4% of the total responses were replaced with the 2 SD value.

Two analyses of variance were conducted on the data, one with subjects ( $F1$ ) and one with items ( $F2$ ) as the random variable. In each analysis, the factors were Group (subject

groups in the subject analysis, item groups in the item analysis), NP head number (singular vs. plural), and Congruence with head (match vs. mismatch). The Group factor was included to extract variance due to the counterbalancing procedure (in the subject analysis, it extracts variance due to differences between the four lists, and in the item analysis, it extracts variance due to differences between the four subject groups). This was a nonrepeated factor in both analyses. The Congruence factor was a repeated measures factor in both analyses, as was the factor of NP head number. The central effect of interest is the interaction between NP head number and Congruence.

Subjects' mean RTs and error rates for the response to the verb are shown in Table 1. The pattern of mean RTs clearly shows that the presence of a mismatch when the head is singular (SP) increases processing time for the verb by 70 ms (relative to the SS control), but there is no mismatch effect at all when the head is plural (PS vs. PP).

As can be seen in Table 1, there is an interaction between NP head number and Congruence. Both analyses of variance show this interaction to be significant ( $F1(1,28) = 5.76, p < .05; F2(1,44) = 5.27, p < .05$ ). A simple comparison of SS vs. SP also shows a significant effect of congruence ( $F1(1,28) = 16.10, p < .001; F2(1,44) = 16.27, p < .001$ ). In addition, there was a significant main effect of Congruence ( $F1(1,28) = 9.41, p < .01; F2(1,44) = 6.33, p < .05$ ). The effect of head number was not significant ( $F(1,28) = 1.77, p < .05; F2(1,44) = 3.38, p > .05$ ).

### *Discussion*

These results support the notion that erroneous feature migration is not purely a production phenomenon: It happens in comprehension, as well.

There is another possibility, however, and that is that the Maze task is not a pure comprehension task, but involves covert production on the part of the participants. That is, it could be argued that the way in which the choice between alternatives is made is by appeal to a sentence construction process. The major

TABLE 1  
 MEAN READING TIMES (IN MS) AND ERROR RATES FOR THE VERB IN THE FOUR SENTENCE  
 TYPES TESTED IN EXPERIMENT 1

Sentence types		RT	% error
SS	The author of the speech is here now.	669	1.04
SP	The author of the speeches is here now.	739	1.30
PP	The authors of the speeches are here now.	724	1.30
PS	The authors of the speech are here now.	724	1.04

Note. Sentence codes (Singular and Plural) refer to head number and local noun number.

similarity between the tasks is that a selection must be made from among a number of lexical options. It is therefore not impossible that participants in this experiment recruited production mechanisms in order to carry out the task.

It is important, therefore, to explore whether the same pattern of results can be found with a task which is more clearly a comprehension task. This is the purpose of the next experiment.

## EXPERIMENT 2

The purpose of this experiment is to verify that the results obtained in Experiment 1 are replicable with a comprehension task that does not require covert production. The task used is the sentence classification task (Forster & Olbrei, 1973), in which participants were required to read a string of words which appeared in full on a computer screen, as in normal text, and to judge whether the words appeared in the proper order. This task is a whole-sentence technique, which does not allow an on-line analysis of momentary processing load. However, for current purposes, the question of where the difficulty arises is not at issue, and hence an off-line measurement of processing time is adequate. In fact, it is advantageous, since it is bound to detect an increase in processing time *no matter when it occurs*. To be sure that the task was sensitive to processing difficulty, we also included a set of items which we expected, based on past research (e.g., Wanner & Maratsos, 1978) to produce reading time differences. This contrast involved subject relatives (such as 11a)

and object relatives (such as 11b). Any failure to detect a difference between these conditions (subject relatives should be faster) would raise doubts about the sensitivity of the experiment.

(11a) Bruce resented the woman who married his father.

(11b) Bruce resented the woman who his father married.

In this experiment, the distracters (the ill-formed sentences) consisted of items such as those in (12) below. Again, it should be noted that this task does not focus attention on the question of subject-verb agreement, since none of the distracters was ungrammatical by virtue of subject-verb disagreement. The ungrammaticality of these items consisted of gross violations of word-order constraints.

## Method

*Participants.* Twenty-nine undergraduates enrolled in an introductory Psychology course at the University of Arizona participated in this study for course credit.

*Materials and procedure.* The experimental stimuli were as in Experiment 1. Sentences were presented in their entirety on a computer screen. Each sentence appeared on the screen until the subject's button press, or for 3 s. Participants were instructed to read each string of words and to decide whether the sequence of words was legitimate in English. Subjects pressed a button as soon as they had decided whether the sentence contained an acceptable sequence of words. Filler sentences contained



TABLE 2  
MEAN RESPONSE TIMES (IN MS) AND ERROR RATES FOR THE FOUR SENTENCE  
CONDITIONS EXAMINED IN EXPERIMENT 2

Sentence types		RT	% error
SS	The author of the speech was subsequently well rewarded.	1920	3.6
SP	The author of the speeches was subsequently well rewarded.	2044	6.8
PP	The authors of the speeches were subsequently well rewarded.	1995	3.9
PS	The authors of the speech were subsequently well rewarded.	1994	3.0

illicit sequences which were blatantly unacceptable, such as the following:

(12a) The car powerful quickly past drove other competitors the.

(12b) The large pumpkin was the for used pie pumpkin.

In addition, 20 pairs of relative clause sentences were included; one variant of each pair was a subject-relative sentence, the other was an object-relative sentence. These items were counterbalanced across presentation lists 1 and 2 and again across lists 3 and 4.

In all, there were 88 grammatical sentences and 88 ungrammatical sentences. These were preceded by 12 practice items.

### Results

The data from one subject were excluded due to a high error rate. Just under 5% of the overall responses were replaced with the 2 SD value.

To be absolutely certain that our task measures relative comprehension difficulty, we first compared the subject relatives vs. object relatives, which had been included as an indicator of task sensitivity. The results were exactly as predicted: The mean reading time for subject relatives was 1902 ms (4.8% errors) and for object relatives was 2036 ms (9.1% errors). Analyses of variance showed that both the RT difference and the error difference were highly significant; for RTs,  $F(1,24) = 32.01, p < .001$ ;  $F(1,76) = 30.83, p < .001$ ; for errors,  $F(1,24) = 8.47, p < .01$ ;  $F(1,76) = 9.35, p < .01$ .

Mean whole-sentence reading times for the four sentence types are presented in Table 2.

The pattern of reading times shown in Table 2 is identical to that found in Experiment 1. There is a striking difference between the SS and SP conditions (124 ms). But there is no difference at all between the PP and PS conditions (-1 ms). Two analyses of variance were conducted, one with subjects as the random variable, the other with items as the random variable. The critical interaction of head-number and congruence was significant ( $F(1,24) = 11.32, p < .01$ ;  $F(2,144) = 6.98, p < .05$ ). In addition, there was a main effect of congruence ( $F(1,24) = 10.37, p < .01$ ;  $F(2,144) = 7.46, p < .01$ ).

Error rates mirrored the RT data; the critical interaction was significant ( $F(1,24) = 3.95, p = .058$ ;  $F(2,144) = 4.97, p < .05$ ).

### Discussion

The comprehension task used here shows the same pattern of reading times as did the Maze task: The SP condition produces significantly longer response times than its congruent counterpart, SS; the PS condition produces no corresponding slowdown. This is consistent with the notion that the plural feature on the local noun may sometimes migrate, creating uncertainty as to whether the singular form of the verb is correct.

It is possible, however, that there are subtle differences in semantic plausibility or complexity among the four conditions that may have contributed to the observed effects, despite the fact that a counterbalancing procedure was used to roughly equate the four conditions for semantic content. It could be, for instance, that the sentences in the SS condition were semantically more

plausible than the SP sentences, and therefore easier to process. For example, in (10a–d), although all conditions refer to the rewarding of speech authorship, it might be the case that an SS concept such as *The author of the speech* is perceived to be more plausible than the corresponding SP concept *The author of the speeches*. However, there may be no corresponding difference in plausibility between *The authors of the speeches* and *The authors of the speech*. Of course, this differential pattern of plausibility would have to hold systematically across items.

Controlling for such effects is extremely difficult, since it is not clear which aspect of plausibility is most relevant to sentence processing (e.g., Forster, 1987), nor is it clear how this aspect should be measured. Fortunately, in the present context, there is a simple way to test whether the experimental effects are due to syntactic processes involving agreement or to accidental confounds with plausibility. The procedure simply involves altering the verbs so that they are uninflected, e.g., *will be rewarded* instead of *is/are rewarded*. This completely eliminates the issue of agreement, and if there are any plausibility or conceptual differences associated with the SP condition, then this effect should still be apparent.

### EXPERIMENT 3

The purpose of this experiment was to test whether the effects observed in Experiments 1 and 2 could have been due to accidental differences in semantic or conceptual plausibility. The experiment used the same materials as in Experiment 1, except that the tense of the verbs was altered so that the verb no longer had to agree in number with the head noun (e.g., *The authors of the speech will be rewarded*). If the previous effects were due to agreement processing, then there should be no differences at all between the SS and SP, or the PP and PS conditions with uninflected verbs. However, if these effects were due to semantic or conceptual factors, then these effects should still be apparent.

### Method

*Participants.* Twenty-nine University of Arizona students participated in the experiment for course credit.

*Materials and procedure.* Experimental items were modified so that all copular forms were replaced with modals, as in the following example:

(13a) SS The author of the speech will be well rewarded.

(13b) SP The author of the speeches will be well rewarded.

(13c) PP The authors of the speeches will be well rewarded.

(13d) PS The authors of the speech will be well rewarded.

In all other respects, the materials and design were the same as in Experiment 2. The procedure was exactly the same as in Experiment 2.

### Results

Data from one subject who made more than 20% errors were excluded from analysis. Just under 5% of the responses were changed to the 2 SD value. As Table 3 shows, the pattern of reading times looks entirely different when the verbs are uninflected for number. Most striking is the fact that the difference between the SS and SP conditions has now disappeared.

Again, two analyses of variance were conducted, one with subjects, the other with items, as the random variable. There were no significant main effects. The interaction of head-number and congruence did not reach significance ( $F(1,24) = 3.39, p > .05$ ;  $F(1,44) = 1.93, p > .05$ ). Pairwise comparisons revealed no significant difference between the SS and SP conditions (both  $F$ 's  $< 1$ ). The PP and PS conditions also did not differ significantly ( $F(1,24) = 4.16, p > .05$ ;  $F(1,44) = 1.93, p > .05$ ), suggesting that the numerical difference between the two plural head conditions was likely due to only a subset of the items.

In addition, these data were compared to

TABLE 3

MEAN READING TIMES (IN MS) AND ERROR RATES FOR THE FOUR SENTENCE CONDITIONS TESTED  
IN EXPERIMENT 3, WHICH USED UNINFLECTED VERBS

Sentence types		RT	% error
SS	The author of the speech will be well rewarded.	2020	3.6
SP	The author of the speeches will be well rewarded.	2035	4.8
PP	The authors of the speeches will be well rewarded.	2033	3.9
PS	The authors of the speech will be well rewarded.	1985	4.2

those obtained in Experiment 2. If we consider the stimuli with uninflected verbs to represent a baseline, we can now verify that the SS – SP difference obtained in Experiment 2 is significantly different from this base line. Analyses of variance show that the SS – SP difference in Experiment 2 is significantly greater than that obtained in this experiment ( $F1(1,48) = 10.68, p < .01$ ;  $F2(1,44) = 6.626, p < .05$ ). In contrast, the difference between the PP and PS conditions is not significantly different in the two experiments ( $F1(1,48) = 1.76, p > .05$ ;  $F2(1,44) = 1.48, p > .05$ ).

Error analysis revealed no significant effects.

### Discussion

Response times for this experiment show a pattern different from that found for Experiments 1 and 2. When the verb is uninflected for number, the SP condition no longer produces significantly slower RTs than the SS condition. This suggests that the slowdown in reading times for the SP condition is indeed linked to subject-verb agreement, and is not due to inherent differences in plausibility or complexity.

Thus far, the comprehension experiments show a mismatch effect like that observed in production—it occurs only with singular heads. But the production studies in English have also shown that conceptual number does not play a role in the implementation of subject-verb agreement. As we described earlier, Bock and Miller (1991) explored whether grammatically singular but notionally plural

NPs induced a greater number of errors than NPs which were both grammatically and notionally singular. They found equivalent error rates for the two types of NPs: Notional number did not create additional errors. Although the implementation of subject-verb agreement in English seems uninfluenced by conceptual factors, agreement computation in comprehension may be affected by the notional number of the subject NP. As observed above, the goals of the production and comprehension systems are distinct. While production processes are occupied with finding and producing the appropriate forms of words and sentences, the comprehension system is devoted to computing an intended message. Semantic integration occurs quickly (e.g., Fodor et al., 1996), and once the meanings of individual words are appropriately combined, syntactic details may not be retained. If the appearance of an inflected verb initiates a check of the number of the subject NP to ensure that there is a number match, it is not at all clear that it must be a syntactic representation that is consulted; rather, such a feature-checking operation could refer to a conceptual representation of the subject NP. If so, there may be a mismatch effect, but only when the subject NP is notionally plural. The next experiment explores this question.

### EXPERIMENT 4

The aim of this experiment was to explore effects of notional number on the comprehension of sentences in which the verb is inflected for number. Since no mismatch effect was observed for the plural head conditions in Exper-

iment 1 and 2, these conditions were omitted. Thus, we focus here on whether the magnitude of the mismatch effect (SP minus SS) varies according to whether the SP condition implies single or multiple tokens.

The materials for this experiment were pre-tested to ensure that readers agreed with the experimenters about whether the NPs implied single or multiple tokens, and in order to determine whether there were any semantic plausibility or complexity differences between them.

### Method

*Participants.* One hundred six students participated in this study for course credit: 32 participated in the pre-test; 30 participated in a plausibility rating study; 44 participated in the comprehension test.

*Materials.* Eight noun phrases of the form *the key to the cabinets* were created. Half of these were intended to be construed as *single-token* phrases (e.g., *The bridge across the canyons*) and half as *multiple-token* phrases (e.g., *The address on the envelopes*). These were then randomized and rated by 32 participants in a pretest. Raters were asked to read each phrase and indicate whether it “was about one thing or more than one thing.” The thirteen best single-token items and 13 best multiple-token items were selected. Within the former group, the range was from 31/32 to 21/32 “one thing” judgments, with a mean of 26.2 (or 81%); within the latter group, the range was from 2/32 to 16/32 such judgments, with a mean of 11.7 (or 37%). For each of these 26 items, the SS counterparts were created.

To ensure that the SS and SP versions of each item were equally plausible, these sentences were also tested in an offline plausibility judgment task. Also included in this test for comparison purposes were the sentences tested in Experiment 2, and an additional twenty implausible sentences (e.g., *The fingerprint of the cakes was found outside*). Subjects were asked to judge, on a 7-point scale, the likelihood and naturalness of the depicted event. (A score of 7 was to be assigned to natural and likely scenarios, a score of 1 to

impossible scenarios). The results of the judgment task showed that the mean ratings for the plausible sentences were all in the 6–7 range, while implausible sentences received a mean rating of 2. The SS and SP versions of the single- and multiple-token items showed virtually no difference: for single tokens, the means were 6.53 and 6.46, a difference of .07, and for multiple-token items, 6.34 and 6.27, again, a difference of .07. These differences were comparable to the difference between the SS and SP sentences from Experiment 2: the means were, respectively, 6.26 and 6.21, a difference of .05. Note that this task was sensitive enough to detect differences: the single- and multiple-token sentences (collapsed across the SS and SP versions) showed a statistically significant difference of .18 (6.49 for single-token sentences vs. 6.31 for multiple-token sentences). In sum, these ratings suggest that the SS and SP variants of the two different sentence types are indeed comparable in terms of plausibility.

For the comprehension task, the SS and SP versions were counterbalanced across two presentation lists, interleaved with filler sentences. In all, there were 52 experimental sentences per list (13 SS single-token items, 13 SP single-token items, 13 SS multiple-token items, and 13 SP multiple-token items), 12 grammatical fillers, and 64 ungrammatical fillers. In addition, there were twelve practice trials.

*Procedure.* The procedure was identical to that used in Experiments 2 and 3.

### Results and Discussion

Prior to analysis, 4.2% of the response times were replaced with the 2 SD value. As Table 4 shows, for both sentence types, the match and mismatch conditions show a comparable difference (55 ms for the multiple-token sentences and 65 for the single-token sentences).

Analyses of variance revealed the following: The critical interaction of congruence and item type (single token vs multiple token) was nonsignificant (both  $F_1$  and  $F_2 < 1$ ). However, there was a significant main effect of

TABLE 4

MEAN RESPONSE TIMES (IN MS) AND ERROR RATES FOR THE MATCH AND MISMATCH CONDITIONS FOR SINGLE TOKEN AND MULTIPLE TOKEN ITEMS, TESTED IN EXPERIMENT 4

Sentence types		RT	% error
Multiple token			
SS	The ad on the billboard was very offensive.	1696	2.5
SP	The ad on the billboards was very offensive.	1751	3.3
Single token			
SS	The bridge across the canyon is safe.	1727	3.9
SP	The bridge across the canyons is safe.	1792	2.8

congruence ( $F(1,42) = 15.37, p < .001$ ;  $F(2,148) = 13.8, p < .001$ ). Analyses of errors revealed no main effects or interactions.

These data replicate the singular-head mismatch effect found in Experiment 2. Although the magnitude of the difference is smaller in this experiment than in Experiment 2 (60 ms vs 124 ms), so too are the mean reading times for the experimental stimuli (1742 vs 1982), and it may well be the case that the size of the mismatch effect depends on the overall reading time.

The absence of an interaction suggests that even though a phrase may be interpreted as notionally plural, this does not lead to *additional* uncertainty about the proper form of the verb. It is possible, of course, that the word order judgment task is insensitive to semantic variables. However, this task has been shown to be sensitive to quite subtle differences in semantic plausibility; faster response times have been reported for sentences such as *The usher took the tickets*, compared to *The tailor took the tickets*, both of which cases are certainly plausible (given the range of meanings for *tickets*) (Forster, 1979).

A question which naturally arises is how deeply participants are processing the sentences in this task. Obviously, the sentences must be processed carefully enough to be able to discriminate the well-formed sentences from the ill-formed sentences. Also, the fact that subject-relatives were easier to process than object relatives in Experiment 2 suggests that the sentences were being subjected to a fairly detailed syntactic analysis. Yet these findings

do not necessitate the construction of a syntactic tree or its equivalent. Given the hypothesis outlined in the introduction that the mismatch effect is due to erroneous feature migration, it would obviously be useful to demonstrate that readers are indeed computing a hierarchical structure of the sentences they are reading.

In the final experiment, we explore erroneous feature migration in pairs of relative clause sentences in which a mismatch appears at different syntactic distances to the head NP, as shown in sentence examples (8b) and (9b), repeated here as (14b) and (15b) (the all-singular controls are given in (14a) and (15a)):

(14a) The owner of the house who charmed the realtor was no longer willing to sell.

(14b) The owner of the house who charmed the realtors was no longer willing to sell.

(15a) The owner of the house which charmed the realtors was no longer willing to sell.

(15b) The owner of the house which charmed the realtors was no longer willing to sell.

In (14), the relative clause modifies the head NP (we will refer to this as the *High Attachment Condition*); in (15), the relative clause modifies the second NP (the *Low Attachment Condition*). As noted above, the linear distance between the head and mismatching NP is the same in both (b) versions.

If erroneous migration of the plural feature creates uncertainty about the correct number of the verb, then there should be greater uncertainty in just that condition in which there is a tighter syntactic link between the mismatch and the head, in the

High Attachment Condition. Obviously, this link between the head and the relative clause must be made in order for differences to emerge. Hence, a superficial analysis of these sentences—arguably, all that is really required by the word order judgment task—will not be sufficient for differences to emerge. Experiment 5 tests just this possibility.

#### EXPERIMENT 5

The purpose of this experiment was to explore the nature of the syntactic representation that readers compute during sentence comprehension. If readers properly compute structures in which a constituent modifies either the head NP or an intermediate NP, then a mismatch effect should be observed for the structure containing a mismatching NP which is syntactically close to the head. As in the previous experiment, we are interested in whether there is an interaction between sentence type and the plural vs singular variants; specifically, whether the mismatch effect is greater for high attachment than low attachment sentences.

#### *Method*

*Participants.* Fifty-three students from the University of Arizona participated in the comprehension study, and 26 participated in the plausibility judgment task. They received course credit for their participation.

*Materials.* Thirty-two quadruplets such as those in (14) and (15) were created. These sentences were subjected to the same plausibility judgment task as were the stimuli in Experiment 4; again, a 7-point scale was used, with 7 representing likely, highly plausible scenarios and 1 representing impossible ones. The quadruplets, and 20 identically structured implausible quadruplets, were counterbalanced across four presentation lists. The mean rating for the experimental items was 6.04 (vs 2.01 for the implausible sentences). Differences in attachment produced very little effect: the high attachment sentences received a mean rating of 6.03 and the low attachment sentences a mean rating of 6.07. The singular

and plural versions showed virtually no difference: for the high attachment cases, the means were 6.03 and 6.02, respectively, and for the low attachment cases, the means were 6.08 and 6.06.

For the comprehension study, the quadruplets were counterbalanced across four presentation lists. In addition, there were 96 other sentences. Of these, 64 contained ungrammatical word sequences, and 32 contained (grammatical) relative clauses, of which sixteen were subject-relative clause constructions, and sixteen were object-relative clause constructions, such as those described earlier in (11a) and (11b) respectively. Again, these were included as an indicator of task sensitivity. If the experimental sentences failed to show a mismatch effect, we would need to ensure that the participants were closely attending to the sentences. The relative clause sentences were counterbalanced across lists 1 and 2 and again across lists 3 and 4. In all, each presentation list contained 32 experimental sentences (eight of each type), 32 relative clause sentences, 64 filler sentences, and eight practice trials.

*Procedure.* The procedure was the same as that used in Experiments 2–4, except that the experimental sentences and an equal number of fillers appeared on two lines. The experimental sentences were divided at different points, but all variants of a given item contained a line break at exactly the same position in the sentence.

#### *Results and Discussion*

Again, the sentences used as a task-sensitivity index—the subject and object relative clauses—showed a robust difference. The subjects' means were (2003 ms) and (2122 ms) respectively. These were significant in both analyses ( $F(1,48) = 31.27, p < .001$ ;  $F(1,60) = 6.11, p < .05$ ).

The subjects' mean response times and error rates in the conditions of interest are presented in Table 5. The means suggest that a slowdown associated with mismatching NPs is apparent only in the high attachment condition, the condition in which there is a more

TABLE 5

MEAN RESPONSE TIMES (IN MS) AND ERROR RATES FOR THE MATCH AND MISMATCH CONDITIONS FOR HIGH VS. LOW ATTACHMENT OF THE MODIFYING CLAUSE, TESTED IN EXPERIMENT 5

Sentence types		RT	% error
High attachment			
SS	The owner of the house who charmed the realtor was no longer willing to sell.	2495	5.6
SP	The owner of the house who charmed the realtors was no longer willing to sell.	2562	6.8
Low attachment			
SS	The owner of the house which charmed the realtor was no longer willing to sell.	2585	5.8
SP	The owner of the house which charmed the realtors was no longer willing to sell.	2536	6.0

direct syntactic link to the head NP from the mismatching NP. Oddly, in the low attachment condition, mean RTs are faster for the SSP versions of sentences than SSS versions.

Analyses of variance reveal the following: Although there were no significant main effects of either attachment site or congruence, the critical interaction of the two factors was found to be significant by both subjects and items ( $F(1,48) = 11.04, p < .01$ ;  $F(1,28) = 10.10, p < .01$ ). Pairwise comparisons show that the difference between the two high attachment conditions is significant by subjects, but only marginally so by items ( $F(1,48) = 7.54, p < .01$ ;  $F(1,28) = 3.97, p = .056$ ). The difference between the low attachment conditions is not significant on either analysis ( $F(1,48) = 2.64, p > .05$ ;  $F(1,28) = 3.08, p > .05$ ). Error rates show little variation, and analyses of variance showed no significant effects or interactions.

The major finding of this experiment was that there was a significant difference between the SS and SP conditions only in cases where the clause containing the mismatching NP is linked directly to the head NP. This suggests that the internal structure of the subject NP has a significant impact on whether a mismatching NP will cause processing disruption. This, in turn, suggests that our participants were fully parsing the sentences.

#### GENERAL DISCUSSION

To summarize, Experiments 1 and 2 examined the comprehension of sentences in which

the complex NP subject contained NPs which matched or mismatched in number with the head NP. Reading times were slower when there was a mismatch, but only when the head noun was *singular*. When the head noun was plural, a singular local noun did not make processing more difficult. Experiment 3 was conducted to ensure that the asymmetric mismatch effect obtained in Experiments 1 and 2 did not arise spuriously, as a result, for example, of inherent differences in complexity or plausibility among the sentence types. Therefore, verbs which were uninflected for number were used. This change produced an entirely different pattern of results; most notably, the significant slowdown observed in the SP condition in Experiments 1 and 2 was not apparent here. This suggests that the mismatch effect was not due to an inadvertent bias in the materials, but rather to the inflectional agreement properties of the sentences. The last two experiments extended the basic finding in two ways. Experiment 4 examined SP cases in which the grammatically singular head noun was, conceptually, either singular (the *single-token* cases) or potentially plural (the *multiple-token* cases). The latter type might be expected to increase uncertainty about the proper number of the head NP, thereby increasing the difference between the SS versions and SP versions. But in fact, the single- and multiple-token cases exhibited an equivalent slowdown in the SP condition. (Numerically, the *single-*

*token* cases showed a larger difference, but this was not statistically significant). In the final experiment (Experiment 5), the syntactic structure of the subject NP was manipulated such that the syntactic distance between the head NP and mismatching NP varied, but linear distance was held constant. The results showed that syntactic distance had an effect: the SS – SP slowdown emerged only when the mismatching NP was linked directly to the head NP.

The results of these experiments reflect the results for production experiments: The conditions which are associated with the production of a greater number of subject–verb agreement errors are just those which are associated with longer reading times.

The results are also compatible with a growing body of research on the comprehension of agreement. We noted in the introduction a number of experiments which showed that a subject–verb mismatch slows down processing in a variety of different tasks. Many of these studies examined the effects of subject–verb agreement *errors* following complex subject NP's containing matching or mismatching nouns, which may focus participants' attention on agreement. Even so, the results of such studies bear on the question of whether the asymmetric mismatch effect found in sentence production studies arises as a direct result of how agreement is implemented in production. If this effect were purely a byproduct of production processes, then it would not turn up in a reading task, however attuned to agreement participants may be. In addition to such studies, however, there are several unpublished reports of experiments which have used, as we did, fully grammatical sentences, and techniques and materials which do not focus attention on subject–verb agreement. These have revealed a mismatch effect for singular but not plural heads. For example, Stevenson (1993) found a mismatch effect with grammatical sentences using a sentence-matching paradigm, in which participants simply decide whether two visually-presented sentences are the same or not. Also, Pearlmutter, Garnsey, and Bock (1995), using both eye-tracking and self-paced reading techniques, have found slower

reading times on the verb (and on the word following the verb) when the complex subject contains a singular head and a mismatching local NP. This result supports at least the inference that the comprehension effects are not limited to the two rather dissimilar techniques used in this paper. Since comprehenders show a mismatch effect in a variety of reading situations, we will assume, tentatively, that this effect was not induced by the particular tasks.

Although it is possible that the mismatch effect in production and comprehension have different sources, the most parsimonious account of these sets of findings is that the same mechanisms are involved in production errors and reading slowdown. Let us consider again the interpretation of production errors. Bock has argued that, in production, the computation of subject–verb agreement is purely syntactic: Conceptual variables do not affect error incidence, but syntactic variables do (Bock, 1995). Quite consistent with this view is the notion that feature migration is syntax-dependent. Vigliocco and Nicol (1996) have sketched out the following mechanism for feature transmission during production. Normally, the number feature associated with the head noun will percolate to the NP node.<sup>3</sup> Number-marking of the verb occurs when this feature is transmitted from the highest NP node to the verb. Thus, correct subject–verb agreement involves the upward movement of a number feature within the structure of the subject NP. An error occurs when there is upward percolation of a number feature from a noun which is *not* the head noun. Further, as discussed earlier, there is a fundamental difference between singular and plural nouns; following Bock and Eberhard (1993), they assume that only the latter is feature-marked,

<sup>3</sup> We assume that although the source of number is the noun, the NP node is ultimately marked, and the verb agrees with this NP. This allows, for example, subject–verb agreement to be characterized as the agreement in number between the head *NP* and the verb, thereby capturing agreement between a subject pronoun and a verb (e.g., *they are*), a bare plural and a verb (e.g., *dogs are*), a conjoined NP and a verb (e.g., *the girl and the boy are*), etc.



and since a feature must be present in order to percolate, only plural local nouns can cause interference. Presumably, both the comprehension and production systems require the construction and representation of syntactic structure in working memory. If the mismatch effect in comprehension and production does have a single cause, it is likely that it is the vulnerability of the structured sentence representation to erroneous feature movement.

This is not to say that subject-verb agreement *per se* is computed during comprehension the same way it is in production. As pointed out above, the comprehension device need not specify the number of an upcoming verb because this would only be useful if the parser could be certain that the upcoming verb would be inflected. Further, if the parser specified verb number based on, for example, the number of the first NP it encountered, an error could result if the first NP turned out to be part of a conjoined NP, since this complex subject NP would require a plural verb (e.g., *the key to the cabinets and the code for the copier are in the desk drawer*). The most sensible way for the sentence comprehension device to operate would be to compute subject-verb agreement in reverse: given an inflected verb, the subject NP (or its memory representation) is checked for number. Erroneous feature percolation would create a subject-verb mismatch and require additional checking. If the verb is not inflected for number, then the head NP will agree in number, and no further checking would be necessary. This would explain why the mismatch effect found in Experiments 1 and 2 did not turn up in Experiment 3. Another difference between the production and comprehension of agreement arises with respect to the structures examined in Experiment 5; in these constructions, the head noun was modified by a PP, which was followed by an embedded clause. In the comprehension experiment, attachment site had an effect. A mismatch effect was observed when the mismatching NP appeared in a constituent which was attached high, directly to the head NP. In the production experiment reported in Nicol (1995), there was no effect of attachment, and nearly no effect of a mismatch. The interpretation of

this was that the mismatching NP and the head were not active at the same time and that the verb number had already been determined by the time the mismatch became active. Now consider comprehension. If the number of the subject NP is checked only once the verb is encountered, then the entire subject NP, including the final constituent and any plural NP within it, will have been processed by the time checking occurs. There still appears to be some limit on interference, since a mismatching NP creates interference only if it is linked directly to the head NP.<sup>4</sup> If the mismatching NP is embedded too deeply, it apparently has no interference effect whatever.

Note that the results from Experiment 5 do not dictate that the representation that is examined during the "backward-looking agreement check" is syntactic; these results, and the results of Experiments 1 and 2, are equally compatible with a *conceptual* number-checking mechanism. Indeed, given the apparent speed with which word meanings are integrated during sentence processing (Swinney, 1979), and the findings for rapid semantic and plausibility influences on syntactic ambiguity resolution (e.g., MacDonald et al, 1994), one might expect that a conceptual representation of the complex subject NP would be checked. However, the results of Experiment 4 suggest that conceptual number does *not* enter into the agreement-checking process. Hence, taken together, the results of these experiments suggest that, in English, agreement checking is a purely syntactic process which is blind to conceptual or semantic factors, just as it appears to be in production.

#### CONCLUSION

To summarize, based on the findings for English, we have proposed that (1) the mismatch effect in both comprehension and production may be due to erroneous feature per-

<sup>4</sup> This finding is at odds with (unpublished) results reported by Pearlmutter, Garnsey, and Bock (1995), who found no such effect of syntactic distance. However, because they used different sentence structures and different techniques, it is difficult to evaluate the source of the discrepancy.

colation within a hierarchically structured representation of a sentence or sentence fragment in working memory; (2) the nature of the computation of subject–verb agreement in production is “forward-specifying,” while in comprehension, it is “backward-checking” (with this directional asymmetry deriving from the processing requirements of the two systems); (3) the computation of subject–verb agreement in both comprehension and production involves syntactic aspects of the subject NP, and not conceptual ones.

A seeming implication of the present study is that the production and comprehension systems show significant overlap in their operations. Initially, this may seem surprising. One might have reasonably assumed that the two processes are so specialized and efficient at solving the problems in their particular domains that their operations have little in common. It is likely that, in the main, this is exactly the right assumption: the computational operations and processing routines are indeed likely to be different. But the working memory representations might be the same, and any error in computing or maintaining the representation is likely to emerge in both comprehension and production. Hence, the overlap may actually be quite small, restricted just to agreement and other discontinuous dependencies which require matching over a potentially unbounded distance, and thus must make reference to a structured memory representation.

#### REFERENCES

- ANTON-MENDEZ, I. (1996). Clitics and attraction errors: An experimental study of language production. Unpublished manuscript, University of Arizona.
- BOCK, K. (1995). Producing agreement. *Current Directions in Psychological Science*, **4**, 56–61.
- BOCK, J. K., & CUTTING, J. C. (1992). Regulating mental energy: Performance units in language production. *Memory and Language*, **31**, 99–127.
- BOCK, J. K., & EBERHARD, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes*, **8**, 57–99.
- BOCK, J. K., & MILLER, C. A. (1991). Broken agreement. *Cognitive Psychology*, **23**, 45–93.
- CUTTING, J. C., NICOL, J. L., & BOCK, J. K. (1995, November). Controlling pronoun number across clause boundaries. Presented at the 36th annual meeting of the Psychonomic Society, Los Angeles, CA.
- EBERHARD, K. M. (1993). *The specification of grammatical number in English*. Unpublished doctoral dissertation, Michigan State University.
- FODOR, J. D., NI, W., CRAIN, S., & SHANKWEILER, D. (1996). Tasks and timing in the perception of linguistic anomaly. *Journal of Psycholinguistic Research*, **25**(1), 25–57.
- FORSTER, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Erlbaum: Hillsdale, NJ.
- FORSTER, K. I. (1987). Binding, plausibility, and modularity. In J. L. Garfield (Ed.), *Modularity in Knowledge Representation and Natural Language Processing*. Bradford-MIT Press: Cambridge, MA (pp. 63–82).
- FORSTER, K. I., & OLBREI, I. (1973). Semantic heuristics and syntactic analysis. *Cognition*, **2**, 319–347.
- FRAZIER, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, **5**, 519–559.
- FREEDMAN, S. E., & FORSTER, K. I. (1985). The psychological status of overgenerated sentences. *Cognition*, **19**, 101–131.
- FRIEDERICI, A. D., & MECKLINGER, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research*, **25**(1), 157–176.
- GOODGLASS, H., & KAPLAN, E. (1972). *The Assessment of Aphasia and Related Disorders*. Philadelphia, PA: Lea & Febiger.
- HAGOORT, P., BROWN, C., & GROOHAUSEN, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, **8**(4), 439–484.
- KEENEY, T. J., & WOLFE, J. (1972). The acquisition of agreement in English. *Journal of Verbal Learning and Verbal Behavior*, **11**, 698–705.
- MACDONALD, M., PEARLMUTTER, N., & SEIDENBERG, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676–703.
- NEVILLE, H., NICOL, J. L., BARSS, A., FORSTER, K. I., & GARRETT, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, **3**(2), 151–165.
- NICOL, J. (1995). Effects of clausal structure on subject–verb agreement errors. *Journal of Psycholinguistic Research*, **24**(6), 507–516.
- NICOL, J., & VIGLIOCCO, G. (1997). Subject–verb misinflections: Interference from within clauses. Manuscript in preparation.
- OSTERHOUT, L., HOLCOMB, P. J., & SWINNEY, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology*, **20**(4), 786–803.
- OSTERHOUT, L., & MOBLEY, L. A. (1995). Event-related

- brain potentials elicited by failure to agree. *Journal of Memory and Language*, **34**, 739–773.
- PEARLMUTTER, N. J., GARNSEY, S. M., & BOCK, J. K. (1995, November). Subject-verb agreement processing in sentence comprehension. Poster presented at the 36th annual meeting of the Psychonomic Society, Los Angeles, CA.
- RATCLIFF, J. (1983). *Inference Processes in the Early Stages of Sentence Comprehension: A Study of the Plausibility Effect*. Unpublished doctoral dissertation, Monash University, Melbourne, Australia.
- SEVALD, C. A., & GARNSEY, S. M. (1995, March). Safe syntax: Encapsulation of number-marking information in sentence comprehension. Presented at the Eighth Annual CUNY Conference on Human Sentence Processing, Tucson, AZ.
- STEVENSON, B. (1993). *Ungrammaticality and stages in sentence processing*. Unpublished doctoral dissertation, Monash University, Melbourne, Australia.
- SWINNEY, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, **18**, 645–660.
- VIGLIOCCO, G., BUTTERWORTH, B., & GARRETT, M. F. (1996). Cross-linguistic variability in subject-verb agreement errors. *Cognition*.
- VIGLIOCCO, G., BUTTERWORTH, B., & SEMENZA, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, **34**, 186–215.
- VIGLIOCCO, G., & NICOL, J. (1997). The role of syntactic tree structure in the construction of subject-verb agreement. Manuscript submitted for publication.
- WANNER, E., & MARATSOS, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 118–161). MIT Press: Cambridge, MA.

(Received February 6, 1996)

(Revised manuscript received July 30, 1996)