ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Memory and Language 48 (2003) 740–759

Journal of Memory and Language

www.elsevier.com/locate/jml

Plausibility and grammatical agreement

Robert Thornton[a,*] and Maryellen C. MacDonald[b]

[a] *University of California, San Diego, USA*
[b] *University of Wisconsin, Madison, USA*

Received 3 June 2002; revision received 12 November 2002

Abstract

Three experiments examined plausibility effects on the production and comprehension of subject–verb agreement. In a production task, participants were given a verb and sentence preamble and asked to create a complete passive sentence. The preambles contained two nouns (e.g., *the album by the classical composers*). The plausibility of the verb was manipulated so that either (a) both nouns could be plausible passive subjects (e.g., *praised*, as both albums and composers can plausibly be praised) or (b) only the head noun could be a plausible subject (e.g., *played*, as only albums can plausibly be played). The comprehension task was self-paced reading with the same materials. The results from both methodologies demonstrated robust plausibility effects. There were higher agreement error rates in production and longer RTs at the verb in comprehension when both nouns were plausible subjects than when only the head was plausible. Implications for current production models are considered and an alternative account is presented that is motivated by current comprehension models and other recent production data.

Many of the world's languages exhibit agreement phenomena, meaning that the forms of two or more words in a sentence match in some way. For example, in English, finite verbs typically agree with their subjects, so that in the sentence *my best friend is here*, both the noun and verb are singular, whereas in the sentence *my best friends are here*, both are plural. Because agreement is a fundamental aspect of language processing, studies of grammatical agreement have played an increasingly central role in language production research in recent years.

To examine how speakers produce agreement, production researchers have used a sentence completion task to elicit subject–verb agreement errors and thereby shed light on the agreement production processes (e.g., Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001; Bock & Miller, 1991; Eberhard, 1997; Vigliocco, Butterworth, & Garrett, 1996). In this task, participants are given a sentence subject and asked to use it in producing a complete sentence. On critical trials, the subject contains two noun phrases (e.g., *the key to the cabinets*), one of which is the *head* NP (*the key*), because it contains the head noun of the subject, and the other is referred to as the *local* NP. Several studies have found that when the number marking on the local NP mismatched that of the head, participants were more likely to produce an agreement error than when the number of both nouns matched (e.g., Bock & Miller, 1991). These findings, along with other studies showing that non-syntactic factors do not seem to modulate agreement error rates (Bock & Miller, 1991; Bock & Eberhard, 1993), led many production researchers to assume that the mechanism underlying agreement errors is some sort of syntactic feature processing during a strictly grammatical stage of the production process (Bock & Eberhard, 1993; Eberhard, 1997). On this view, which we will term the

* Corresponding author. Present address: Department of Linguistics and Cognitive Science, 550 Harvard Ave., Pomona College, Claremont, CA 91711, USA. Fax: 1-909-621-8065.
E-mail address: robert.thornton@pomona.edu (R. Thornton).

0749-596X/03/$ - see front matter
doi:10.1016/S0749-596X(03)00003-2

encapsulated model, production processes are separated into several distinct processing stages (e.g., Bock & Levelt, 1994; Garrett, 1980; Levelt, Roelofs, & Meyer, 1999, although cf. Vigliocco & Hartsuiker, 2002). In the encapsulated model, processing begins with the formulation of a non-linguistic conceptual representation. Then, through *grammatical encoding*, each lexical concept is mapped onto its abstract syntactic representation. On this view, agreement processes occur during grammatical encoding and non-syntactic information is not thought to affect this process. Semantic and conceptual information affect message formulation and even the mapping from the message to the grammatical stage of production, but agreement operates over purely syntactic representations. Data taken as support for the encapsulated model have typically demonstrated dissociations among the use of semantic, syntactic, and phonological information, and this is exactly what was found in the earlier agreement studies (Bock & Eberhard, 1993; Bock & Miller, 1991).

More recently, though, several agreement studies have found semantic effects on agreement error rates under certain conditions. For example, studies by Vigliocco and colleagues examined semantic constraints on agreement through the use of phrases that have distributive readings, such as *the label on the bottles*. The heads of such phrases (e.g., *label* in the example), although grammatically singular, are often interpreted as referring to multiple labels, one on each bottle. Vigliocco et al. (1996) found that such phrases elicited more agreement errors than control phrases in Spanish, but not English. Thus, their study supported Bock and colleagues' findings that semantic variables do not affect agreement processes in English, but suggested that semantic variables may play a role in other languages. These findings have been further complicated by a study by Eberhard (1999), who demonstrated semantic effects in English, but only when other factors were tightly controlled. Even though these results challenged the interpretation of earlier findings, the semantic effects have been somewhat inconsistent and have generally been incorporated within the encapsulated model (e.g., Bock et al., 2001). Thus, the encapsulated model maintains the assumptions of principles and parameters approaches in linguistics (e.g., Chomsky, 1981), namely that agreement is primarily a syntactic phenomenon: a redundancy in morphosyntactic features accomplished through feature copying or coindexation. Limited semantic effects are explained as arising in the message level, affecting the mapping to the grammatical level, but not affecting agreement processes directly (Bock et al., 2001).

Other linguistics approaches, however, provide support for the view that agreement is a referential, rather than syntactic, phenomenon. Barlow (1999) and Pollard and Sag (1988) provide cross-linguistic evidence that agreement can be systematically determined on the basis of discourse information, even when that discourse information conflicts with number marking. For example, the sentences in (1) are identical except for the verb, which is singular in (1a) and plural in (1b).

(1) a. My best friend and harshest critic is here.

b. My best friend and harshest critic are here.

Either could be grammatically correct, depending on if *my best friend and harshest critic* refers to one or two people. Thus, this is a case in which agreement must be computed on the basis of intended reference, rather than from strict number marking. Similarly, in (2):

(2) Scrambled eggs and biscuits is my favorite breakfast.

The verb is singular, even though the subject *scrambled eggs and biscuits* is plural. Again, this is done for semantic reasons, as the scrambled eggs and biscuits are treated as a single entity: a breakfast.

The fact that agreement is in some cases determined by discourse factors suggests that non-syntactic constraints may play a larger role in agreement than is currently assumed in the encapsulated model. The purpose of the current paper is to pursue this possibility. We examine the influence of non-syntactic constraints in language production by examining plausibility effects on subject–verb agreement in English.

Parallels with comprehension

Although clear examples of referential effects on agreement, as in (1) and (2), exist in the linguistics literature, psycholinguistic evidence for such discourse constraints on agreement production is scarce. However, there are interesting parallels concerning studies examining semantic influences on comprehension.

An influential model of sentence comprehension, Frazier and colleagues' garden-path model, assumes distinct processing stages similar to the encapsulated production model: an initial stage in which an encapsulated syntactic representation is formed, followed by a stage of semantic integration (Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983). Thus, both the garden-path comprehension account and the encapsulated production model fundamentally assume that an independent syntactic representation is needed to mediate the mapping between form and meaning. As in the production literature, comprehension results demonstrating the absence of semantic effects early in comprehension were taken as initial evidence for an independent, encapsulated stage of syntactic processing in comprehension.

This claim of a lack of early semantic effects in comprehension was later questioned by researchers working in a constraint-based framework, who proposed that comprehension involves the early integration

of multiple probabilistic sources of information (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). On this view, the non-syntactic effects on parsing that were previously attributed to later stages of comprehension were instead seen as emerging from the nature of constraint interaction and the speed at which different constraints can affect the comprehension process. For example, research in this framework demonstrated that the effects of subtle semantic or discourse constraints can be obscured by stronger lexical or structural biases, explaining why semantic effects were observed in some studies, but not in others (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald et al., 1994; Spivey & Tanenhaus, 1998). The equivocal data concerning semantic effects on agreement (e.g., Eberhard, 1999; Vigliocco et al., 1996) may reflect a similar situation: the subtle semantic or conceptual manipulations that have been employed to date may have had minimal effects on agreement production as compared to the stronger effects of number marking. This explanation implicates a potentially much larger role of semantic variables in production models, which might interact with syntactic and morphological factors.

Borrowing from the work in comprehension, if we consider production from a constraint-based perspective, computing agreement can be conceptualized as a process in which multiple sources of information are integrated in producing an inflected verb form. On this account, factors should modulate agreement processes to the extent that they have been reliably correlated with the use of a particular verb form. As a result, not all factors necessarily exert an equal effect, so that more reliable constraints may exert a stronger bias, possibly obscuring the effects of more subtle ones. For example, the grammatical number of a noun seems to be a strong cue to verb agreement, as evidenced by the fact that singular nouns typically take singular verbs, and plural nouns take plural verbs. But grammatical number is not the sole determinant of agreement; conceptual number also exerts some influence, so that speakers sometimes utter phrases such as *the crew of sailors are*, even though the head noun *crew* is grammatically singular (Bock, Nicol, & Cutting, 1999; Haskell & MacDonald, in press). Although conceptual number seems to have some influence, Bock et al. (1999) demonstrated that grammatical number exerts a much stronger influence on subject–verb agreement, suggesting that it is a dominant constraint that can obscure the effects of weaker factors.

Because some of the most successful semantic manipulations in the comprehension literature have varied the plausibility of noun–verb relationships (Garnsey et al., 1997; Pickering & Traxler, 1998), the current experiments examined noun–verb plausibility effects on agreement. Both a production task and a comprehension task were used to compare the extent to which non-syntactic factors influence agreement in production versus comprehension. In the production task, participants were presented with sentence preambles that contained two NPs, such as *the album by the classical composer*, along with verbs. The plausibility relations were manipulated via the verbs. For one set of verbs, both NPs were plausible subjects of a passive sentence; this is the *Both NPs* condition. For the example, the verb was *praised*, as both albums and composers can plausibly be themes of a praising event. For the other set of verbs, the head NP was still a plausible subject, but the local NP was implausible. This is the *Head Only* condition, and in the example, the verb was *played*, as albums, but not composers, can plausibly be played. A similar manipulation was used by Hupet, Fayol, and Schelstraete (1998), who examined plausibility effects on written production in French.[1] Although they found robust plausibility effects of the sort that we are predicting, there are several aspects of their study that limit our ability to interpret the data. First, the Hupet et al. study was in French, and there are some suggestions that agreement in romance languages may be more influenced by non-syntactic factors than is the case in English (Vigliocco et al., 1996). More importantly, however, the Hupet et al. study used a dictation task, in which participants simply transcribed auditorily presented sentences. Thus, it is not clear whether their task tapped into central language production processes or simply more general memory processes. To address this potential concern, the current study uses the type of fragment completion task widely used in agreement production studies.

These sorts of plausibility effects have not typically been considered within the encapsulated model. Instead, production studies examining semantic effects have focused on the intrinsic semantics features of individual noun phrases, such as animacy or conceptual number (e.g., Barker, Nicol, & Garrett, 2001; Bock, Loebell, & Morey, 1992; Vigliocco et al., 1996). Barker et al. (2001) also incorporated a preliminary investigation of plausibility through a post hoc analysis of plausibility variations in their stimulus materials, but they did not manipulate plausibility directly. In the present studies, however, noun–verb plausibility was explicitly manipulated, so that the noun phrases were held constant across conditions, and plausibility varied as a function of their relationship with the verb. Thus, plausibility is a probabilistic, contingent semantic variable, making it a classic example of non-syntactic information.

If production of an inflected verb form is the result of the integration of multiple soft constraints, then any reliable cue to verb inflection should modulate agree-

[1] We thank Robert Hartsuiker for pointing out this paper to us.

ment processes. Because grammatical subjects will most often be reliably plausible subjects, plausibility might serve as a consistent cue to verb inflection. Thus, we predict that in the Head Only condition, the reduced plausibility of the local NP will minimize possible interference from a mismatching number cue. In contrast, we predict that in the Both NPs condition, the increased plausibility of the local NP would serve to make a mismatching number cue more prominent, increasing agreement error rates relative to the Head Only condition. Conversely, if agreement is controlled only by the grammatical or conceptual number of the subject NP (Bock et al., 2001), then no effects of plausibility are expected. This is because our plausibility manipulation is contained in the verb, so that each level of the plausibility manipulation has the same preamble and thus the same grammatical and conceptual number.

Experiment 1

The purpose of Experiment 1 was to assess the influence of plausibility information on the production of subject–verb agreement errors in English. Participants were presented with a modified version of the sentence completion task used by Bock and Miller (1991). In this task, participants are given a sentence preamble, usually a noun phrase, which serves as the subject of the sentence. Participants then utter a complete sentence, beginning with the preamble and followed by whatever verb and ending comes to mind. In the modified version employed here, participants were given the verb to be produced as well as the preamble. On a given trial, participants were presented visually with a verb, displayed on a computer screen, which was followed by the auditory presentation of a sentence preamble. For example, a trial might consist of the word PRAISED appearing on the screen, followed by the auditory presentation of the preamble "the album by the classical composers." The task was to utter a complete passive sentence out loud, beginning with the preamble and followed by the passive form of the verb and whatever ending first came to mind, for example, "the album by the classical composers was praised by the radio station." The main dependent variable was the proportion of agreement errors of the auxiliary verb, which was usually *was* or *were*.

The advantage of this modified version of the task is that because participants were given both the preamble and verb to be produced, the plausibility relations between the NPs and verb could be tightly controlled. Another methodological advantage is that because we had participants produce passive sentences, almost all of their utterances included a verb clearly marked for number, as passives usually take a form of the verb *be* as an auxiliary. In contrast, the rate of ambiguously number-marked verbs has been relatively high in some of the previous agreement production studies. For example, 57% of the responses in Bock and Miller's Experiment 3 contained an ambiguously number-marked verb.

The specific plausibility manipulation was such that in the Both NPs condition, both the head and local NPs would make relatively plausible subjects of the passive form of the verb and in the Head Only condition, the head NP would make a plausible subject, whereas the local NP was relatively implausible. The other two manipulations involved the grammatical number of the NPs. Both the head and local NPs were presented in both their singular and plural forms, resulting in four number conditions: singular–singular (SS), singular–plural (SP), plural–singular (PS), and plural–plural (PP). A sample item is presented in (3):

(3) a. Both NPs plausible: PRAISED
"the album(s) by the classical composer(s)"
b. Head Only plausible: PLAYED
"the album(s) by the classical composer(s)"

The specific hypothesis being tested was that plausibility is a cue, among others, to which inflected form of the verb should be produced. If so, then error rates should be relatively low in the Head Only condition, because plausibility would favor only the head NP, making the conflicting cue from the local NP less salient.

We do not expect to see plausibility effects in either of the number matched conditions (SS or PP) for two reasons. First, error rates are usually quite low in these conditions, so any effects might be obscured because error rates are at floor, as suggested by Haskell and MacDonald (in press). Second, because the number marking on both the head and local nouns converge on the same verb form, increasing the local noun's plausibility would only provide more support for producing the correct verb form.

In regard to predictions in the number mismatch conditions, in most studies, error rates are significantly lower in the PS than the SP condition. In fact, error rates for the PS condition are typically on par with the PP match condition, suggesting a relative insensitivity to variables that increase agreement error rates. Thus, the influence of plausibility might not be observed in the PS condition because of that insensitivity.

Previous experiments (e.g., Bock & Miller, 1991) have demonstrated that the SP condition typically exhibits the highest proportion of agreement errors, as compared to the few errors in the other number conditions. In the current study, this pattern of data would be reflected in higher error proportions in the SP condition than in the other conditions. We further predict that plausibility effects will be most evident in the SP condition, because error rates in this condition are significantly higher, thus avoiding floor effects and offering the best chance of observing more subtle plausibility effects.

Method

Participants

A total of 190 University of Southern California students participated for extra-credit in undergraduate Psychology courses. Of those, 150 completed the plausibility ratings task and the other 40 participated in the online production task. None were involved with both tasks and all reported that they were native English speakers.

Materials and design

The materials consisted of the 40 sets of sentence preambles and verbs listed in the Appendix. There were two levels of each of the three factors, resulting in eight items in each stimulus set.

Each preamble was of the form NP_{head} *prep* NP_{local}. Half of the head NPs were plural and half were singular. Half of the local NPs were plural and half were singular. For half of the verbs, both the head and local NPs were plausible subjects (Both NPs condition), and for the other half, the head NP was plausible and the local NP was implausible (Head Only condition). A set of two verbs was used for two items. Thus, the verbs associated with the example item *the album by the classical composer*, *praised* and *played*, were also used with the item *the announcement by the factory manager*. These items were counterbalanced across presentation list so that each verb appeared only once on each list.

Sixty filler preambles and verbs were also constructed. The filler preambles consisted of simple NPs. Half of these contained only a determiner and a noun (e.g., *the children*) and the other half contained a determiner and a noun, plus some prenominal modification (e.g., *the chicken burrito*). Half of the nouns were singular and the other half were plural. Thus, half of all materials required a plural agreeing verb and half required a singular verb. None of the nouns or verbs from the fillers were used in any of the experimental items. Five additional items of the same type were also constructed for use as practice items.

Eight counterbalanced stimulus lists were created. Each list contained exactly one of the eight members of each stimulus set, with each member appearing on exactly one list. An equal number of items from each condition appeared on each list. Thus, Plausibility, Head NP Number, and Local NP Number were all within-participant variables. Each list was comprised of 100 items, 40 experimental and 60 filler. The items on each list were presented in random order with the constraint that no two experimental items were adjacent.

Plausibility ratings

Plausibility ratings were collected in order to get an independent, quantifiable measure of the robustness of the plausibility manipulation. A two-page, forty-item survey was constructed for this purpose, which was included in a larger survey packet that was distributed to several undergraduate psychology courses as extra credit.

The surveys were constructed in the following manner. The head and local NPs from each item were separately paired with each of the verbs for that item in a counterbalanced fashion. An example of the singular ratings items from the preamble *the album by the classical composer* is presented in (4):

(4) a. the album was praised
b. the album was played
c. the classical composer was praised
d. the classical composer was played

There were eight conditions per item, with two levels of each of the following three factors: Which NP (Head versus Local), Number (Singular versus Plural), and Plausibility (Both NPs versus Head Only). Each item was followed by a seven-point Likert-type scale. Eight counterbalanced lists were created. Each list had five items from each condition and each item appeared only once on each list. Thus, all three factors, Plausibility, Number, Which NP, were within-participant variables. The items on each list were presented in a different random order and two practice items were provided. Each survey was printed on one double-sided sheet of paper. Participants were instructed to rate how plausible each item was by circling a number on the seven-point scale (for which 1 = very implausible, 4 = neutral, and 7 = very plausible).

The data are illustrated in Fig. 1 and the result of the related ANOVA analyses are presented in Table 1.

These analyses revealed that as predicted, the interaction of Which NP and Plausibility was significant, such that for the head NPs, the Head Only condition was rated as more plausible than the Both NPs condition (6.11 versus 5.24), $F_1(1,149) = 295.33$, $p < .001$; $F_2(1,39) = 22.95$, $p < .001$, whereas for the local NPs, the Both NPs condition was rated as significantly more plausible than the Head Only condition (5.61 versus 2.52), $F_1(1,149) = 1307.84$, $p < .001$; $F_2(1,39) = 343.48$, $p < .001$.

The results demonstrate a much larger difference in plausibility in the Head Only condition than in the Both NPs condition, but there was still a plausibility difference in the Both NPs condition. This effect was such that the local NPs were rated as slightly more plausible than the head NPs even though both were rated as relatively plausible. That withstanding, in order to operationalize plausibility with respect to these ratings, an item was considered plausible if its mean rating was significantly

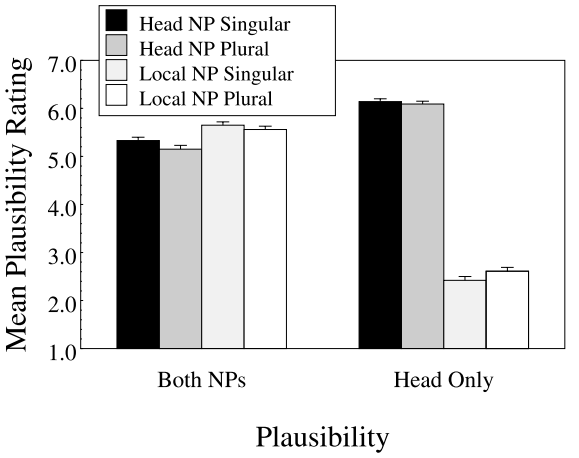

Fig. 1. Plausibility ratings by condition from Experiment 1, with standard error bars computed across participants.

Table 1
Results from ANOVAs on the plausibility ratings data from Experiment 1

Condition	F_1	F_2
Which NP	1004.14*	155.00*
Number	1.03	0.51
Plausibility	595.69*	89.08*
Which NP × Number	6.18*	4.51*
Which NP × Plausibility	1361.98*	229.25*
Number × Plausibility	9.35*	5.78*
Three-way interaction	1.36	0.65

Note. Degrees of freedom for all F_1s $= (1, 149)$ and for F_2s $= (1, 39)$.
* $p < .05$.

greater than 4, the midpoint and neutral value of the scale. An item was considered implausible if its value was significantly less than 4. Single-sample t tests (with $df = 1, 149$) were used to determine if the values differed significantly from the midpoint. All of the items were rated significantly as plausible, all ts > 14.00, all ps $< .001$, except for the local NPs, both singular and plural, in the Head Only condition, which were rated as significantly implausible, ts > 17.00, ps $< .001$. These results indicate that the plausibility manipulation was robust: in the Both NPs condition, both NPs were plausible, whereas in the Head Only condition, only the head was plausible.

As can be seen in Table 1, there were a number of small but reliable effects in addition to the much larger plausibility manipulation. It was not anticipated that these small effects would have any noticeable effect on error rates and so they will not be discussed further here.

Procedure: Production task

The procedure was a variation of the sentence-fragment completion task used by Bock and Miller (1991). Participants were seated in front of a computer and received instructions telling them that they would see a verb displayed on the computer screen, then hear the beginning of a sentence. They were told that shortly after the end of the auditory presentation, the word *Go* would appear on the screen and that this would be their cue to utter a complete passive sentence beginning with the preamble they heard, followed by the appropriate form of the verb they saw and the first ending that came to mind. The on-screen instructions gave an example of an appropriate passive response to an sample item. If participants failed to produce passives on any of the practice trials, they were corrected and given explicit instruction on how to produced a passive (i.e., the verb has to be proceeded with *was* or *were*).

The experiment was run on an Apple Macintosh, using the Psyscope software package and button-box (Cohen, MacWhinney, Flatt, & Provost, 1993) to control presentation. External speakers were attached to the computer in order to present the auditory materials. At the beginning of each trial, a fixation cross appeared at the center of the screen for 500 ms, followed by the visual presentation of the verb. After 1000 ms, the preamble was played from the speaker. The verb remained on the screen for the duration of the preamble, at which time it disappeared. One second after the end of the preamble, the word *Go* appeared on the screen, which was the cue to the participant to initiate their response. The 1 s delay between the offset of the verb and the *Go* cue was included because in a pilot study, participants began responding on some trials before the preamble had finished playing. The delay in the current experiment was intended to allow participants to fully process the preamble on all trials before initiating a response. The materials were presented in two experimental blocks. The first block consisted of the practice items. The second block consisted of the items from one of the stimulus lists. The entire session was recorded on audio tape and took approximately 25 min.

Scoring

The participants' responses were transcribed and placed in one of four categories: *correct responses*, *agreement errors*, *voice errors*, and *preamble errors*. Correct responses required that (a) the preamble was repeated correctly, (b) the response contained the correctly inflected passive form of the verb, and (c) the response was a full sentence uttered in its entirety. The preamble was considered correct if the nouns and

determiners were repeated in the correct order and with the correct number marking. Preambles were considered correct even if the prenominal adjective from the local NP was omitted.

Agreement errors were scored in the same way as correct responses, except that a participant used the incorrect number marking on the auxiliary verb. Cases in which participants initially produced the incorrect form of the auxiliary but then corrected themselves were counted as agreement errors.

Voice errors were responses for which the preamble and verb were repeated correctly and were grammatically correct, but which used a verb voice or tense that did not differentiate between singular and plural agreement. For example, participants sometimes used the active rather than passive voice (e.g., "the protest by the activists interrupted the convention").

Preamble errors were responses for which participants were unable to repeat the preamble or verb correctly. These errors were further classified into four subcategories. (1) *Missed verb* errors included trials in which participants used a different verb or were unable to remember the verb. (2) *Head NP number* errors were responses for which participants changed the number marking on head NP, but repeated the rest of the preamble correctly. (3) *Local NP number* errors were responses for which participants changed the number marking on local NP, but repeated the rest of the preamble correctly. (4) *General* preamble errors included any other preamble errors, which did not fall within the first three subcategories.

Results

Overall, 79.8% of the experimental items were correct, 6.7% were agreement errors, 1.4% were voice errors, and 12.1% were preamble errors. See Table 2 for a full breakdown of response types by condition.

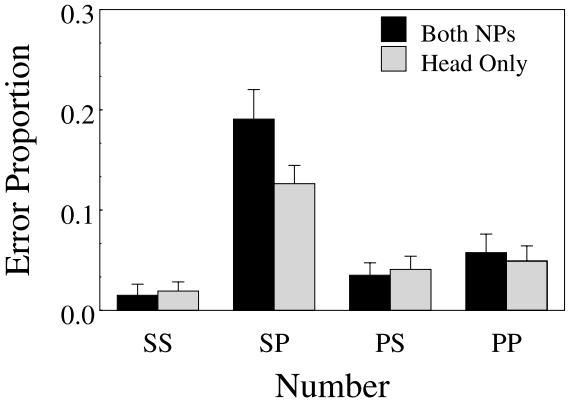

Fig. 2. Mean proportions of subject–verb agreement errors by condition from Experiment 1, with standard error bars computed across participants. SS, singular–singular; SP, singular–plural; PS, plural–singular; PP, plural–plural.

Agreement error analysis

The proportions of agreement errors by condition are presented in Fig. 2 and the results from the related ANOVAs are presented in Table 3.

If plausibility constrains agreement processes, then agreement errors should be lower in the Head Only condition, where plausibility information disfavors incorrect agreement with the local noun, as compared to the Both NPs condition, in which the plausibility constraint is more equally balanced. Moreover, any such effect should be the strongest in the SP condition because of the mismatching number cue and its relatively high error rate, thus offering the best opportunity to observe the potentially subtle effect of plausibility. Because the a priori prediction was that plausibility effects would be greatest in the SP condition, planned pairwise comparisons of Plausibility were conducted in each of the Number conditions.

Table 2
Response breakdown by type and condition from Experiment 1

Response type	Condition								Total
	Both NPs				Head Only				
	SS	SP	PS	PP	SS	SP	PS	PP	
Correct	177	134	151	170	175	153	146	170	1276
Agreement errors	3	39	7	11	4	25	8	10	107
Voice errors	2	3	3	2	2	4	5	2	23
Preamble errors									
Missed verb	6	2	6	4	8	3	10	4	43
Head NP number	1	11	5	6	3	5	8	6	45
Local NP number	10	6	23	3	6	6	17	5	76
General	1	5	5	4	2	4	6	3	30

As predicted, there was a significant plausibility effect in the SP condition, $F_1(1,39) = 5.83$, $p < .05$; $F_2(1,39) = 5.14$, $p < .05$, such that the Both NPs condition yielded significantly more agreement errors than the Head Only condition. There were no significant effects of plausibility in any of the other number conditions, all Fs < 1.

As shown in Table 3, the three-way interaction of Head Number, Local Number, and Plausibility was not significant. This is not surprising, because the predicted pattern of results was that very small or null effects would be observed in three of the four number conditions, with the plausibility effects only predicted in one of the number conditions. Thus, the likely reason that the three-way interaction is not significant is that the plausibility effect in the SP condition is obscured by the lack of any significant differences in the other 75% of the items. We return to this point in Experiment 2.

We also conducted a correlational analysis in order to assess the relationship between our offline plausibility ratings and the agreement errors from our online production task. Our experimental hypothesis was that when the plausibility of both the head and local NPs were relatively high, participants would produce a higher proportion of agreement errors than when only the head NP was plausible. In order to quantify the plausibility differential between the head and local NPs from the rating norms, we used a measure of the relative plausibility advantage of the head NP over the local NP, which we will term the Head Plausibility Advantage (HPA). It was calculated, for each level of the plausibility manipulation for each item, by subtracting the mean plausibility rating for the local NP from the rating for the head NP. Thus, when the HPA is positive, it indicates that the head NP was rated as more plausible than the local NP, and when it was negative, it indicates that the local NP was rated as more plausible than the head NP. And, of course, greater values in either direction indicate a larger differential between the plausibility of the NPs, whereas values closer to zero indicate less of a differential. Because the ratings scale varied from 1 to 7, HPA values could range from −6 to 6. For the Both NPs condition, the HPA range was from −3.22 to 2.21, with a mean of −0.38. For the Head Only condition, the range was from 0.78 to 5.00, with a mean of 3.54.

Table 3
Results for ANOVAs on effects of all three variables and their interactions on agreement errors from Experiment 1

Condition	F_1	F_2
Head NP Number	5.32*	9.53*
Local NP Number	33.43*	35.46*
Plausibility	2.63	1.94
Head × Local	13.35*	35.41*
Head × Plausibility	3.52	1.68
Local × Plausibility	3.63	2.63
Three-way interaction	2.62	1.66

Note. Degrees of freedom for all F_1s $= (1,39)$ and for F_2s $= (1,39)$.
* $p < .05$.

Because the plausibility effect on agreement errors was observed only in the SP condition, we correlated the HPA values with the proportion of agreement errors in the SP condition. We found a marginally significant negative correlation between the two measures, $r(80) = -.19$, $p = .09$. The nature of this trend was that when the plausibility advantage of the head NP increased, the proportion of agreement errors decreased. One potential explanation for the marginal significance of the correlation is that there were a number of items for which participants made very few, if any, agreement errors in the SP condition, even though there was variation in the ratings data for these items. Thus, differences that would be predicted on the basis of the ratings data might be obscured because the agreement error data are at floor for these items. To assess this potential problem, we recomputed the correlation excluding all cells with zero means. This resulted in a stronger, significant correlation between the HPA values and the proportion of agreement errors, $r(46) = -.38$, $p < .01$. Again, the nature of this correlation was that as the plausibility advantage for the head NP increased, the proportion of agreement errors decreased.

Preamble and voice error analyses

In analyzing the preamble errors, ANOVAs revealed that there was a main effect of Local NP Number on the pattern of missed verb errors $F_1(1,39) = 7.08$, $p < .05$; $F_2(1,39) = 7.12$, $p < .05$. The nature of this effect was that participants forgot the verb more frequently when the local NP was singular than when it was plural (3.8% versus 1.6%). This finding is interesting, in that aspects of the preamble affected the likelihood of retaining the verb, suggesting interplay in keeping both in memory.

There were also significant differences in the distribution of local NP number errors. There was a significant main effect of Head NP Number on local NP errors, $F_1(1,39) = 4.88$, $p < .05$; $F_2(1,39) = 6.37$, $p < .05$, such that participants were more likely to change the number of the local NP when the head NP was plural than when it was singular (6.0% versus 3.5%). There was also a significant main effect of Local NP Number, $F_1(1,39) = 19.26$, $p < .05$; $F_2(1,39) = 4.78$, $p < .05$, such that participants were more likely to change the number of the local NP when it was singular than when it was plural (7.0% versus 2.5%). These main effects were qualified by a significant interaction of both variables, $F_1(1,39) = 7.18$, $p < .05$; $F_2(1,39) = 8.66$, $p < .005$. The nature of this interaction was that par-

ticipants were more likely to change the number of the local NP for the PS preambles (10.0%) than for the other ones (rates between 2% and 4%). In other words, the local NP number errors were driven by cases in which participants were making the PS preamble a PP one, thereby making a number mismatched item a matched one. There were no significant effects of plausibility on local NP number errors.

There were no significant differences in the distribution of head NP number errors or general preamble errors across conditions. There were also no significant differences in the distribution of voice errors across condition.

Discussion

There are several important aspects of the data from Experiment 1. First, consistent with our predictions, there was a significant effect of plausibility on the production of agreement errors. In the SP condition, there were significantly more agreement errors for the Both NPs items than for the Head Only ones. Thus, these data support the hypothesis that plausibility is a reliable cue to inflection. When both NPs were relatively plausible and conflicted in number, participants produced significantly more agreement errors. In addition, the distribution of agreement errors in Experiment 1 was asymmetrical across number conditions, with a greater proportion of agreement errors in the SP mismatch condition than in the PS one, replicating the results of previous studies (e.g., Bock & Miller, 1991).

In Experiment 1, we also observed a negative correlation between our plausibility ratings data and the proportion of agreement errors in the SP condition. This relationship was such that as the plausibility advantage for the head NP over the local NP increased, the likelihood of agreement errors in the SP condition decreased, which is consistent with our hypothesis that the increased relative plausibility of the head NP makes the conflicting number cue from the local NP less salient. It should be noted, however, that this finding is inconsistent with a recent study by Barker et al. (2001), who demonstrated reliable effects of two semantic variables, animacy and the semantic overlap between the nouns in the preamble, but found only a small, negative correlation ($r = -.05$) between plausibility and agreement errors. There are several important differences to note between that study and the current one. First, Barker et al. (2001) did not actually manipulate plausibility, but collected post hoc ratings data for a study that had already been conducted. Given the typical strategy in stimulus design to avoid large variations in plausibility across items, it is likely that the plausibility differences within their items were smaller than in the current study, which could result in a weaker correlation. Second, Barker et al. (2001) presented participants with an adjective and preamble, whereas we presented participants with a verb and preamble. It is certainly possible that plausibility manipulations related to verbs could exert a more robust effect on verb agreement that manipulations of adjectives.

Taken together, the results of Experiment 1 confirm the findings of Eberhard (1999) and Barker et al. (2001) that even in English, non-syntactic factors significantly moderate the production of agreement. Moreover, our plausibility results go beyond the earlier findings of lexical semantic effects within the NP to demonstrate the robust effects of a contingent, probabilistic semantic factor on agreement.

However, one potential problem with interpreting the current results is that our production task also had a substantial comprehension component, in that participants first had to comprehend the preambles and verbs. Preamble comprehension is generally not thought to undermine agreement results because there is no comprehension component of the agreeing element itself in the task. Our inclusion of a verb, however, increases the comprehension component of the task. Specifically, the fact that we presented the verb before the preambles might have lead to greater conceptual activation of that verb, and consequently affected the way in which the preambles were interpreted. This conceptual activation of a verb's event semantics likely approximates aspects of naturalistic production, in that speakers presumably have an active conceptual representation of an event before they begin to produce an utterance about that event. However, it limits our ability to compare our results to previous ones, because most of the previous agreement studies presented only a preamble, without any material from the verb phrase, and those studies that have presented additional material all presented adjectives rather than verbs (Barker et al., 2001; Haskell & MacDonald, in press; Vigliocco & Nicol, 1998). A second potential problem with the methodology of Experiment 1 was that there was a 1 s delay between the presentation of the preamble and the cue for participants to respond. The inclusion of this delay was intended to allow participants to fully process the preamble before initiating a response, but it may have increased the comprehension component of the task, and also increased the task difficulty by forcing participants to retain the items in memory. The purpose of Experiment 2 was to address these concerns.

Experiment 2

The methodological differences between Experiment 1 and previous agreement production studies potentially limits our ability to compare our results with existing data. In Experiment 2, we modified the experimental

task in order to minimize these differences. First, we presented the verb after the preamble was played, in order to minimize the activation of event semantics before the presentation of the preambles. Second, we eliminated the 1 s delay between stimulus presentation and a participant's response.

In addition to these methodological changes, we also included only the SS and SP number conditions from Experiment 1. This was done for two reasons. First, the plausibility effects from Experiment 1 were significant only in the SP condition; we can therefore evaluate potential concerns about the task by using only the SP items, along with the SS control condition. Second, by eliminating the PS and PP number conditions, our experimental power was increased from 5 to 10 observations per cell. We hypothesized that the three-way interaction in Experiment 1 was non-significant because null effects were expected in three of the four number conditions, and low power may have limited our ability to detect such an interaction. Using only two number conditions should increase our ability to detect the interaction by increasing power and by cutting the number of cells in which null effects are expected.

If the results of Experiment 1 reflect production processes involved in computing agreement, then changing aspects of how the materials were presented should not have a substantial effect on the general pattern of results. Thus, our predictions are the same as for Experiment 1: a significant plausibility difference in the SP, but not SS, number condition, such that more agreement errors are observed for the Both NPs than for the Head Only items. However, if the results of Experiment 1 were due to contamination from the comprehension components of the task, then changing those aspects of presentation should reduce or eliminate plausibility effects.

Method

Participants

Thirty-four University of Wisconsin students participated for extra-credit in undergraduate Psychology courses. Two participants were excluded for failing to produce passive sentences for the majority of items. These exclusions resulted in 32 total participants in Experiment 2. All participants reported that they were native English speakers.

Materials and design

The 40 sets of sentence preambles and verbs, as well as the fillers and practice items, from Experiment 1 were used. In the current experiment, however, only the SS and SP number conditions were included. This reduction in conditions yielded two levels of the two factors (Local NP Number and Plausibility), resulting in four conditions. These items were counterbalanced across four presentation lists so that each preamble and verb appeared only once per list and an equal number of items from each condition were included in each list. As in Experiment 1, each list was comprised of a total of 100 items: 40 experimental and 60 filler. The items on each list were presented in random order with the constraint that no two experimental items were adjacent. The same practice items from Experiment 1 were used.

Procedure

The procedure for Experiment 2 was the same as in Experiment 1, with two changes. First, in Experiment 2, the auditory preamble was presented before the verb. At the offset of the preamble, the verb for that trial appeared on the computer screen and remained there until the participant began to speak. Thus, the second, related change from Experiment 1 was that participants did not have to wait 1 second after stimulus offset to respond, but could initiate a response as quickly after the presentation of the verb as they were able.

As in Experiment 1, the materials were presented in two experimental blocks. The first block consisted of the practice items. The second block consisted of the items from one of the stimulus lists. The entire session was recorded on audio tape and took approximately 20 min.

Results and discussion

The responses from Experiment 2 were scored in the same way as the responses from Experiment 1. Overall, 85.0% of the experimental items were correct, 4.8% were agreement errors, 3.5% were voice errors, and 6.8% were preamble errors. See Table 4 for a full breakdown of response types by condition.

Agreement error analysis

The proportions of agreement errors by condition are presented in Fig. 3. The results of Experiment 2 replicate the general pattern of data from Experiment 1. There was a significant main effect of number, $F_1(1,31) = 45.00$, $p < .001$; $F_2(1,39) = 60.37$, $p < .001$, such that more agreement errors were produced in the SP than SS number condition. We also observed a main effect of plausibility, which was significant by participants, $F_1(1,31) = 5.91$, $p < .05$, but only marginally so by items, $F_2(1,39) = 3.97$, $p = .053$. The nature of this effect was that more agreement errors were produced in the Both NPs than Head Only condition.

These main effects were qualified by the significant interaction of both variables, $F_1(1,31) = 7.15$, $p < .05$;

Table 4
Response breakdown by type and condition from Experiment 2

Response type	Condition				Total
	Both NPs		Head Only		
	SS	SP	SS	SP	
Correct	295	243	291	261	1090
Agreement errors	0	37	1	23	61
Voice errors	8	14	12	11	45
Preamble errors					
Missed verb	5	3	4	7	19
Head NP number	1	2	1	9	13
Local NP number	9	14	7	5	35
General	3	7	5	5	20

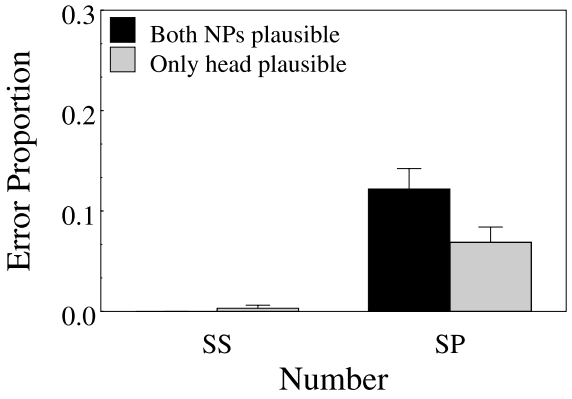

Fig. 3. Mean proportions of subject–verb agreement errors by condition from Experiment 2, with standard error bars computed across participants. SS, singular–singular; SP, singular–plural.

$F_2(1,39) = 5.20$, $p < .05$. Planned pairwise comparisons were also computed, which revealed the nature of this interaction. In the SP number condition, there was a significant effect of plausibility, $F_1(1,31) = 6.67$, $p < .05$; $F_2(1,39) = 4.64$, $p < .05$, such that more agreement errors were produced in the Both NPs condition than in the Head Only condition. In the SS number conditions, there was no significant difference between the plausibility conditions. This significant interaction supports our speculations about why the three-way interaction from Experiment 1 was not significant. In the current experiment, we cut the total number of conditions from eight to four, thereby increasing our power and decreasing the number of conditions with expected null results. Thus, the results of Experiment 2 demonstrate that plausibility and number do interact.

Other errors

In addition to conducting analyses of agreement errors, we also analyzed the other types of errors participants made. There were no differences in the distribution of voice and preamble errors, with the exception of some marginal effects on head number errors. For example, there were more head number errors for the SP preambles in the Both NPs condition than in the Head Only condition, $F_1(1,31) = 8.87$, $p < .01$; $F_2(1,39) = 3.02$, $p = .09$. There was no reliable plausibility difference for head number errors in the SS number condition.

Comparisons with Experiment 1

In order to more directly compare the pattern of results from Experiment 1 to Experiment 2, we conducted an ANOVA with Experiment as a between-participants variable. Because the current experiment only included the SS and SP number conditions, only those items from Experiment 1 were included. Thus, the analysis had three factors: Experiment (1 versus 2), Number (SS versus SP), and Plausibility (Both NPs versus Head Only). The ANOVA revealed a significant interaction of Plausibility and Number, $F_1(1,70) = 10.31$, $p < .005$; $F_2(1,39) = 7.20$, $p < .05$. Pairwise comparisons revealed that the nature of this interaction was that there was a significant difference between the plausibility conditions in the SP condition, $F_1(1,70) = 10.92$, $p < .001$, $F_1(1,39) = 8.16$, $p < .01$, but not in the SS condition.

As expected, there were also main effects of both factors in the combined data. There was a main effect of Number, $F_1(1,70) = 54.45$, $p < .001$; $F_2(1,39) = 99.62$, $p < .001$, such that participants made fewer agreement errors in the SS than SP condition (0.1% versus 4.8%). There was a main effect of Plausibility, $F_1(1,70) = 9.07$, $p < .005$; $F_2(1,39) = 5.98$, $p < .05$, such that participants made more agreement errors in the Both NPs condition than the Head Only condition (8.5% versus 5.6%).

Finally, there was also a main effect of Experiment, $F_1(1,70) = 4.41$, $p < .05$; $F_2(1,39) = 13.78$, $p < .001$, such that participants made more agreement errors on the SS and SP items in Experiment 1 than Experiment 2 (8.9% versus 4.8%). This difference could stem from several factors, including the lack of a response delay in

Experiment 2 versus Experiment 1, and a different subject sample in Wisconsin versus California. Whatever the cause of the overall error rate variation, it is clear that the effect of plausibility was the same in the two studies, as demonstrated by the lack of an interaction of Experiment with Number or Plausibility. Thus, the plausibility effect from Experiment 1 does not seem to be a result of the increased comprehension component relative to previous studies.

The overall pattern of agreement errors from Experiment 2 nicely matches the pattern from Experiment 1. Despite the methodological changes to minimize concerns about the task, we found plausibility effects on agreement production with two different variations of the fragment completion task. There is a wide variety of fragment completion paradigms in the literature, all of which have yielded agreement errors. Fragment completion, like many production methods, does have a comprehension component, but the robustness of the error data over a variety of methods makes it clear that such effects arise during production and cannot be attributed to comprehension. This argument is bolstered by the fact that agreement errors do appear in spontaneous speech (Bock & Miller, 1991) and can be elicited by pictorial, rather than linguistic, stimuli (Haskell & MacDonald, 2002).

Taken together, the results of Experiments 1 and 2 demonstrate that agreement production is sensitive to plausibility, a factor that has been central in constraint-based models of comprehension. Thus, our data suggest potential overlap between production and comprehension processes. In order to directly examine this overlap, we now turn to a comprehension experiment that used the items from the first two experiments.

Experiment 3

In Experiment 3, a word-by-word self-paced reading task was used to present the experimental items from Experiments 1 and 2. Complete sentences were constructed from those items. Each sentence consisted of the sentence preamble, followed by the grammatically correct passive form of the verb and a final *by*-phrase. In the reading task, the stimuli were presented on a computer screen, and participants pressed a key to get each successive word in the sentence. The computer recorded the time between each key press, yielding a measure of the time spent reading each word. Each word was presented individually, with the exception of the auxiliary and verb, which were presented together. They appeared together so that participants would process the verb, and hence the plausibility information, as they were reading the agreeing auxiliary.

Reading times at the verb and surrounding regions were analyzed for effects of the plausibility and number manipulations. In the first two experiments, the Both NPs condition resulted in higher error proportions than in the Head Only condition. Specifically, the locus of the plausibility effect from the first two experiments was in the SP condition. If agreement is the result of similar constraint-satisfaction processes in both production and comprehension, plausibility effects in comprehension should mirror the data from Experiments 1 and 2. This pattern would be reflected by increased reading times in the SP condition at the verb and possibly over subsequent words, with longer reading times in the Both NPs condition, where plausibility is more equally weighed, than in the Head Only condition, where plausibility supports the head NP and thus the appropriate verb form.

Alternatively, it is possible that subject–verb agreement is not as salient in comprehension as in production. Although comprehension studies have yielded results consistent with the extant production data (Nicol, Forster, & Veres, 1997; Pearlmutter, 2000; Pearlmutter, Garnsey, & Bock, 1999), other recent studies have suggested that comprehenders do not necessarily compute full linguistic representations (Christianson, Hollingworth, Halliwell, & Ferreira, 2001). As a result, comprehenders may not fully process agreement information, as it is not strictly essential to comprehending many sentences. If plausibility does not reliably affect the comprehension of agreement, there should be no difference in reading time between the Both NPs and Head Only conditions.

Method

Participants

Twenty-four University of Southern California students participated for extra-credit in undergraduate Psychology courses. All of them reported that they were native English speakers. An additional four participants were run, but excluded from subsequent analyses either because their overall mean reading time was more than 2 SD away from the grand mean for all participants or for missing more than 20% of the comprehension questions.

Materials and design

Forty sets of complete, grammatically appropriate sentences were created from the stimulus sets from Experiment 1. A complete list of these items is presented in the Appendix. The complete sentences were constructed by inserting the appropriate auxiliary (either *was* or *were*) between the preamble and the verb. Endings were also added, which consisted of a *by*-phrase containing a definite NP, such as *by the radio station*.

Each new stimulus set had eight versions, with the same conditions as in Experiment 1: Head NP Number (Singular versus Plural), Local NP Number (Singular versus Plural), and Plausibility (Both NPs versus Head Only). As with Experiment 1, these manipulations yielded two match conditions, singular–singular (SS) and plural–plural (PP), and two mismatch conditions, singular–plural (SP) and plural–singular (PS).

Sixty-four filler items were also created, many of which were adapted from the filler items from Experiments 1 and 2. The fillers were all complete sentences, beginning with a simple NP consisting of the definite determiner *the*, a noun, and in some cases an adjective as well. Half of the NPs were singular and the other half were plural. These NPs were followed by passive verb forms, which used the auxiliary *was* for the singular items and *were* for the plural items. All fillers ended with an agentive *by*-phrase for the passive verb. As in the first two experiments, none of the nouns or verbs from the fillers were used in any of the experimental items. Nine additional items of the same type were also constructed for use as practice items.

In order to insure that participants were reading for comprehension, yes/no comprehension questions were composed for each of the experimental, filler, and practice items. For example, the question related to the item *the album by the classical composer was played by the radio station* was *was it a TV station that was involved with the composer?*. Half of the correct answers to the questions were *yes* and the other half were *no*.

As in Experiment 1, eight counterbalanced stimulus lists were created. Each list contained exactly one of the eight members of each stimulus set, with each member appearing on exactly one list. An equal number of items from each condition appeared on each list. Each list was comprised of a total of 113 items: 40 experimental, 64 filler, and 9 practice. Each list began with the same nine practice items, but the ordering of fillers and experimental sentences was randomized for each participant.

Procedure

The materials were presented on a computer screen using a single-word self-paced reading task. At the beginning of each trial, a line of dashes appeared on the screen, with each dash representing a character from the current item. Participants were instructed to press the space bar of the computer's keyboard to see each word of the sentence in a non-cumulative fashion (Just, Carpenter, & Woolley, 1982). The keypress that ended the presentation of the last word of a sentence triggered the presentation of the comprehension question. Participants answered the question by pressing a key marked either *yes* or *no* and were given feedback on screen about their accuracy. Participants were run individually in a quiet room at the University of Southern California. On average, the experimental session took less than 25 min.

Results

Comprehension accuracy

Overall participants answered 94.0% ($SD = 3.3\%$) of the comprehension questions correctly, all participants answered at least 80% correctly, and only items for which the question was answered correctly were included in analyses. Analyses of comprehension question accuracy revealed that accuracy varied significantly across Head NP Number conditions, such that participants were more accurate when the head NP was singular than when it was plural, 94% versus 90%, $F_1(1,23) = 5.23$, $p < .05$; $F_2(1,39) = 5.61$, $p < .05$. Interestingly, this finding mirrors the plausibility ratings data from Experiment 1, in which the singular head NP items were rated as more plausible than the plural ones. Accuracy did not vary significantly with any of the other conditions.

Reading time analysis

A length-adjusted residual reading time was computed for each sentence region (see Ferreira & Clifton, 1986; Trueswell et al., 1994) by calculating a regression equation across all experimental and filler items, by participant, which found the best linear fit between region length and reading time. The adjusted reading times were the difference between the predicted reading time for a particular region and the actual reading time for that region. Thus, positive adjusted reading times reflected slower than expected processing, whereas negative adjusted reading times reflected faster than expected processing. The motivation for this conversion was twofold: (1) to make comparisons across conditions that varied in number of characters possible and (2) to reduce variance across participants due to differences in overall reading speed and sensitivity to variation in word length.

A two-pass trimming method was used to exclude outliers. First, before calculating this regression equation, all raw RTs more than 2500 ms or less than 100 ms were coded as equipment failures and excluded from analysis. This affected less than 0.3% of all observations (96/40436). Then, all length-adjusted reading times that were more than two standard deviations from the mean were trimmed for each word in each condition, which affected less than 5% of all observations.

For the purposes of analysis, reading times were examined from four critical sentence regions: (1) the noun from the local NP, the last region before the plausibility manipulation is revealed, (2) the auxiliary and verb, where the plausibility effects were predicted to be observed, (3) the word *by*, and (4) the determiner, *the*, from the NP in the *by*-phrase. Reading times for these critical regions are shown in Fig. 4.

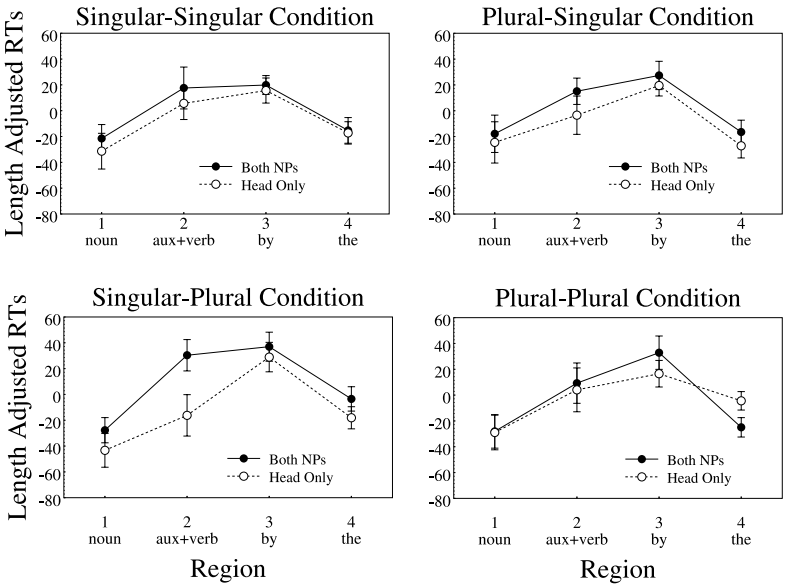

Fig. 4. Mean length-adjusted reading times by condition and region from Experiment 3, with standard error bars computed across participants.

As in the first two experiments, the a priori prediction was that plausibility effects would be greatest in the SP number condition. Moreover, the plausibility effects were specifically predicted to appear in Region 2, which contained the auxiliary and verb. To test this prediction, planned pairwise comparisons were conducted and revealed that in Region 2, there was a significant effect of plausibility in the SP condition, in the predicted direction, $F_1(1,23) = 12.19$, $p < .005$; $F_2(1,39) = 4.96$, $p < .05$. Reading times were significantly shorter for the Head Only condition than the Both NPs condition, consistent with the predictions. There were no significant plausibility effect in any of the other number conditions for this region.

As in Experiment 1, we computed the correlation between online processing difficulty, reading times in the SP condition at the verb region, and rated HPA (head plausibility advantage). We found a significant negative correlation, $r(80) = -.23$, $p < .05$. The nature of this effect was that as the plausibility advantage for the head NP over the local NP increased, reading times at the verb decreased. Thus, this relationship mirrors the correlation from Experiment 1: as the rated HPA increased, processing difficulty decreased.

Discussion

The results of Experiment 3 demonstrated a robust effect of plausibility on the comprehension of subject–verb agreement. The nature of the plausibility effect was that processing was more difficult when both of the NPs were plausible subjects than when only the head NP was plausible. In the current experiment, this difficulty was reflected in longer reading times at the verb in the Both NPs condition than in the Head Only condition. Thus, the comprehension data from the current experiment overlap substantially with the production results from the first two experiments. Not only were there significant plausibility effects in the same direction for both methodologies, but the effects were isolated to the same condition; in all of the experiments, the plausibility effect was observed only in the SP mismatch condition. Taken together, the results from all three experiments demonstrate that plausibility significantly affects both the production and comprehension of agreement in a similar fashion.

The results of Experiment 3 also make a methodological point, that it is possible to obtain interpretable reading data from a paradigm that combines single-word self-paced reading with the occasional presentation of a multi-word phrase. The clear pattern of data from the current experiment indicates that at least limited mixed presentation is feasible when circumstances demand it.

General discussion

The experiments presented here yielded two central findings. First, in Experiments 1 and 2, the plausibility of an NP as a potential subject of a verb significantly affected the production of subject–verb agreement. In these studies, when an NP that was a plausible subject intervened between the head NP and its verb, agreement

processing was more difficult than when the intervening NP was an implausible subject, resulting in a higher proportion of agreement errors. Second, we found analogous plausibility effects on agreement comprehension in Experiment 3. In this case, processing difficulty was reflected by increased reading times at the verb. Moreover, the plausibility effects were strongly parallel across production and comprehension, in that the effects were in the same direction and localized in the SP number condition in all of the experiments. Thus, the current set of results supports the hypothesis that plausibility constrains agreement processes in both production and comprehension in a similar manner. We consider this point in further detail below, first in an exploration of implications for accounts of agreement production and then with respect to production–comprehension overlap more generally.

Non-syntactic effects in agreement production

As noted in the introduction, a strictly encapsulated model of agreement production does not permit the influence of non-syntactic information in the execution of syntactic production processes. Whereas early work supported this approach (Bock & Eberhard, 1993; Bock & Miller, 1991), results from more recent studies have motivated some modifications of this view, such that properties of the message can number-mark a constituent directly. Importantly, these approaches limit semantic influences to properties of the subject NP, such as conceptual plurality of the head noun (Bock et al., 2001). By contrast, our plausibility manipulation involved information outside of the subject NP. Thus our results may suggest that the limited amounts of semantic involvement postulated by Bock et al. do not go far enough in incorporating non-syntactic information into the agreement process.

Before we pursue this interpretation, however, we consider an alternative account of our findings that would preserve the limited effects of semantic information on agreement postulated by Bock et al. (2001). This alternative, which was invoked by Hupet et al. (1998) to account for plausibility effects on agreement in a dictation task in French, posits two distinct processes contributing to the agreement errors that we observed. The first of these occurs prior to agreement processing, when the subject of a sentence is initially selected, and the second is agreement processing itself, where errors can arise from incorrect number feature percolation (e.g., because of number mismatch between a head and local noun). In this alternative account, our plausibility manipulation affects only the early subject selection process, and not agreement processing itself. That is, the increased plausibility of the local noun in our Both NPs condition could result in accidentally selecting this local noun as the subject NP on some trials. If agreement then proceeded correctly, the verb would agree with the local rather than head NP. Thus, what appeared to be a plausibility-based agreement error could instead be a plausibility-based subject selection error, followed by correct agreement with the wrong head noun.

There are two problems with this alternative account. First, in the encapsulated model, subject selection occurs before the linear order of the utterance has been determined (Bock & Levelt, 1994). Consequently, if the local NP was misselected as the subject, then it should have been placed in the (syntactic) subject position and occurred as the first NP in the utterance. This did not happen on any of the trials in the current experiments, and all agreement errors reported here came from trials in which the putatively misselected subject noun was in fact produced correctly in the local noun position, not in the subject position. Thus these claims are inconsistent with results concerning the timing of grammatical role selection and linear ordering operations (e.g., Bock & Levelt, 1994). Second, this alternative account predicts a different pattern of data than was observed in the current experiment. The alternative account posits an effect of plausibility on subject selection, followed by a later independent effect of number attraction on agreement. This should yield a pair of main effects: more errors in the Both NPs condition than in the Head Only condition (across both PS and SP conditions) and a main effect of head versus local plural marking, with more errors in SP than PS conditions (across both levels of plausibility), but crucially no interaction between the factors. However, this is not the pattern of data we observed, as there was no effect of plausibility in the PS condition. Rather than evidence for separate effects of the variables, the interaction we observed in Experiment 2 indicates that plausibility and grammatical number affected the agreement process at the same time, otherwise they would not have been able to interact. Although additional work is needed in these areas, particularly in the process of subject selection, our results are most consistent with the hypothesis that plausibility affects agreement processes directly, although it may affect subject selection as well.

Additionally, our plausibility account is consistent with recent work in both linguistics and psycholinguistics. For example, plausibility effects on agreement are very compatible with linguistic accounts that afford a prominent role for discourse-level information in agreement processes (e.g., Barlow, 1999; Pollard & Sag, 1988). Similarly, a broader role for semantic factors in agreement processing has been increasingly postulated in the language production literature (Barker et al., 2001; Eberhard, 1999; Solomon & Pearlmutter, 2002; Vigliocco et al., 1996; Vigliocco & Franck, 1999). The major contrast of scope between these previous studies and the current work is that previous research has been focused on the internal properties of noun phrases, in

that it has investigated lexical semantic properties of individual nouns (Barker et al., 2001; Vigliocco & Franck, 1999) or relationships between nouns within an NP (Eberhard, 1999; Solomon & Pearlmutter, 2002; Vigliocco et al., 1996). Along these lines, Hartsuiker, Anton-Mendez, and van Zee (2001) noted that there seems to be "a (tacit) assumption in some of the literature that only information contained within the subject phrase can interfere with the production of subject verb agreement" (p. 549). In contrast, we have focused on the relationship between the NP and its agreeing verb. All of the conditions in our plausibility manipulation had the same preamble and thus the same grammatical and conceptual number but they differed in the plausibility relationship between the nouns in the preamble and the verb. Our work therefore suggests that not only do NP-internal semantic factors modulate agreement processes, but that NP-external ones can as well. This demonstration that factors outside of the subject NP can affect agreement also complements the results of Hartsuiker et al. (2001), who found that number mismatching nouns in a direct object, rather than subject, NP increased agreement error rates in Dutch.

Probabilistic constraints in production

We have suggested that agreement processing is guided by the application of multiple probabilistic constraints, and the current results support the existence of noun–verb plausibility relationships as one of those constraints. Although agreement processes have not typically been characterized within a constraint-based framework, there are definite precursors in the production literature. For example, Dell and colleagues (e.g., Dell, 1986, 1990; Dell, Chang, & Griffin, 1999; Dell, Reed, Adams, & Meyer, 2000) have developed a framework for production research that focuses on the interaction of multiple sources of information. This framework affords a prominent role for the influence of distributional information, the statistical or probabilistic knowledge of the order and patterning of linguistic information (e.g., phonemes, syllables, words, phrases) in the input, acquired through experience. Specifically related is a study by Dell et al. (2000), which demonstrated that the types of speech errors participants made were strongly constrained by even very recent distributional information regarding phonotactics.

Other recent data support the substantial role of distributional information in production. For example, Bock and Griffin (2000) offered evidence of long-term structural priming, the finding that speakers often reuse grammatical constructions that have been recently used, even if an alternative construction is permissible. Bock and Griffin suggested that their results are better understood in terms of implicit learning, rather than more traditional spreading activation accounts of priming. This is a considerable departure in thinking from the encapsulated model, which has usually been cast in interactive spreading activation terms (e.g., Pickering, Branigan, Cleland, & Stewart, 2000). Although Bock and Griffin (2000) do not explicitly make this point, their account overlaps considerably with constraint-based models of comprehension. Central to both are a sensitivity to distributional statistics and the notion that the processing of a particular construction is crucially tied to previous experience with it. Similarly, another source of evidence for the influence of distributional information on production is offered by Stallings, MacDonald, and O'Seaghdha (1998), who examined factors involved in the heavy-NP shift construction in English. Heavy-NP shift is the tendency for speakers to shift relatively long direct-object NPs to the ends of sentences from their canonical post-verbal position. One of Stallings et al.'s central findings was that the tendency to shift was lexically specific. Direct objects were more or less likely to shift depending on the verb that headed the verb phrase, and that preference was related to a verb's distributional history of appearing with non-adjacent complements.

With increasing support for the influence of these types of probabilistic constraints in production, it becomes useful to address how a constraint-based model would specifically account for the current plausibility results. Within this framework, computing agreement is not the copying of the number feature from the subject NP onto the verb, but a constraint-satisfaction process in which multiple cues are integrated in the production of an inflected verb form. Plausibility information is one such constraint and can be viewed as a distributional factor analogous to the distributions of speech sounds that have been shown to modulate phonological speech errors (Dell et al., 2000). In the case of plausibility, the distributional information is that the agreeing noun and verb will tend to have a closer plausibility relationship than this verb will have with other nouns. Thus over the history of prior utterances, plausibility information will come to exert some influence over the computation of a verb's inflection. In related work, we have also begun to explore the extent to which distributional information may underlie the SP asymmetry (Thornton, Haskell, & MacDonald, 2001), which is the finding that a disproportionate number of agreement errors occur when a mismatching local noun is plural rather than singular.

Although this account of agreement processing is currently lacking in detail or implementation, it is broadly consistent with many connectionist models. For example, simple recurrent networks (SRNs) have been used to model sentence processing phenomena, including agreement processing (Christiansen & Chater, 1999; Elman, 1991). In a production version of such a model,

the semantic representation of a verb would be mapped onto its inflected form. Presumably, when there are number mismatched NPs in the input, computing agreement would be more difficult because there would be conflicting cues as to which inflected form should be produced. Such models are sensitive to distributional cues in the input, including plausibility. For example, although not applied to agreement phenomena, Harm, Thornton, and MacDonald (2000) recently developed a model that demonstrated such contingent plausibility effects in comprehension. This model was able to calculate distributional statistics over a distributed semantic representation of each word in the input string. It subsequently developed a representation of the contexts that a word appears in, yielding a plausibility measure of that word appearing in certain events. Because the model generated this measure of the plausible semantics of possible continuations, it began to partially activate the relevant semantic features of the upcoming word before it was presented, such that plausible continuations (i.e., words with consistent semantic features) were easier to process. Thus, in this model, plausibility effects arose because at given point in processing, the current input reliably cued semantic features of the subsequent input (see Federmeier & Kutas, 1999; Schwanenflugel & Shoben, 1985, for empirical support for such models). The results of the experiments presented here are readily interpretable within this type of framework. Distributional information regarding the plausibility of the nouns that have been produced significantly influenced the production of agreement at the verb. Distributionally, verbs tend to be more closely related to head nouns than to other nouns in the sentence, and the production system can use this probabilistic information during agreement production. Thus, the same production mechanisms that have been implicated in structural priming (Bock & Griffin, 2000) and sensitivity to phonotactics (Dell et al., 2000) appear here in sensitivity to plausibility information. Our proposal is clearly less specific than the syntactic feature processing accounts that have been developed over a longer period in the literature, but our intention is for it to serve as a framework for discussion of how plausibility effects and other seemingly distinct production phenomena might be linked.

Production–comprehension overlap

The substantial overlap between our production and comprehension data suggests that similar mechanisms may underlie agreement processes in both domains. Our comprehension results confirm previous studies that found additional comprehension difficulty in SP conditions (Nicol et al., 1997; Pearlmutter, 2000; Pearlmutter et al., 1999) and extend them to show that this difficulty can be modulated by plausibility information. These results are also consistent with recent work in structural priming showing that choice of syntactic structure in production can be modulated by the structures that have been recently comprehended (e.g., Bock, 2002; Branigan, Pickering, & Cleland, 2000). Thus there is a growing body of work suggesting that shared representations and similar processing mechanisms may underlie both comprehension and production.

Given these trends, overlap between production and comprehension processes is an important topic for additional research, but it does not appear likely that overlap to the degree obtained in our studies will always be observed. It is important to keep in mind reasons why production and comprehension data might differ in many circumstances. Though they may be affected in the same way by similar information, production and comprehension are fundamentally different tasks, and the behavior of each will likely vary as a function of differing task demands. For example, numerous types of ambiguity need to be resolved in comprehension (see Frazier & Clifton, 1996; MacDonald et al., 1994), but ambiguity resolution is less of a concern in production, because speakers already know the meaning of what they are producing, potentially minimizing the influence of linguistic ambiguity in the production system.

The influence of differing task demands can also be seen in several pilot experiments for the present study, which we think are important in interpreting production–comprehension overlap. The pilot studies used items and tasks similar to Experiments 1–3, but differed in several important ways, all of which contributed to an increased memory component in the production task. First, in the pilot experiments the local NPs were several words longer. For example, an item contained *the internationally respected classical composer* as a local noun, rather than the item that used *the classical composer* in the current study. Second, in the pilot production task, the verb disappeared from the screen before the preamble was played, whereas in the current task the verb remained on the screen. Both of these differences led to an increased memory component in the pilot production task, making it more difficult than the task from Experiments 1 and 2. This difficulty was reflected in the overall rate of preamble errors: 27.2% in the pilot experiment compared to the 12.1% rate in Experiment 1. The pilot production results yielded significant plausibility effects, but in the PP as well as the SP number condition. This finding suggested that the increased task difficulty resulted in increased error rates, even in one of the number matched conditions. This finding is generally consistent with the results of Fayol, Largy, and Lemaire (1994), who found that greater memory load increased agreement errors rates in written production in French. Our pilot compre-

hension experiment used items adapted from the pilot production study, but the longer preambles did not increase the memory load as much because the items did not have to be held in memory for later production. Consequently, the pilot comprehension experiment yielded the same pattern of data as in Experiment 3: significant plausibility effects, but only in the SP condition. Thus, substantial overlap was observed between the pilot production and comprehension experiments, in that both demonstrated significant plausibility effects, but that overlap was not as complete as that between Experiments 1–2 and Experiment 3. Along with other recent findings that have illustrated the importance of considering task demands in interpreting production data (e.g., Kello, Plaut, & MacWhinney, 2000), our current and pilot data indicate that the observed overlap between the production and comprehension data is, to some degree, a function of the difficulty of the task.

In sum, the current results demonstrate that plausibility significantly mediates agreement processes in both production and comprehension, suggesting that similar mechanisms are at work in both domains. In the comprehension literature, the existence of effects of this sort, together with computational accounts of constraint interaction, have resulted in the increased prominence of models without strictly encapsulated stages of processing. In these accounts, the typically different strengths and timecourses of syntactic and non-syntactic constraints are thought to emerge from the interactions of distributional patterns in the language and the computational mechanisms that weigh the constraints. Although it is much to soon to tell whether results of the sort presented here will similarly promote production models with less distinct stages, we do suggest that constraint-satisfaction is a fruitful framework for examining agreement and other production phenomena (e.g., Bock & Griffin, 2000; Dell et al., 2000; Haskell & MacDonald, in press; Stallings et al., 1998; Vigliocco & Hartsuiker, 2002).

Acknowledgments

This research was supported by a University of Southern California University Dissertation Fellowship to the first author and NSF Grant SBR-9511270 to the second author. The experiments presented here were conducted in partial fulfillment of the first author's doctoral dissertation. We thank Todd Haskell, Jack Hawkins, Toby Mintz, Mark Seidenberg, Andrea Thornton, Robert Hartsuikier, and three anonymous reviewers for comments and Chris Hines, Jessica Manela, Julie Moffitt, Allison Neslage, and Gary Oppenheim for help with participant running.

Appendix: Experimental stimuli

For each item, the first verb was the Both NPs verb and the second was the Head Only verb, and only one of them was presented with the item on each trial. For Experiments 1 and 2, which used production tasks, the verb was presented first, followed by the preamble, which is underlined below. For Experiment 3, which used a comprehension task, the entire item was presented as a complete sentence.

Although only the singular–singular number condition is presented in this section, both of the nouns in each preamble were presented in both their singular and plural forms, resulting in four number conditions per item in Experiments 1 and 3. Thus, because both of the verbs for an item were paired with each of the number conditions, there were eight versions of each item in those experiments. In Experiment 2, only the number of the local noun was manipulated, resulting in four conditions. Although only the singular version of the auxiliary (i.e., *was*) is presented below, in Experiment 3, when the head noun was plural, the grammatically appropriate auxiliary was presented (i.e., *were*). Also note that each pair of verbs was used for two items. Which verb was presented was counterbalanced so that a verb appeared only once on each experimental list.

1. The feeling about the undergraduate student was noticed/shared by the dean.
2. The idea of the prominent psychologist was noticed/shared by the hospital.
3. The report about the democratic senator was discovered/published during the election.
4. The picture of the jewel thief was discovered/published by the FBI.
5. The party for the fraternity brother was photographed/catered by the house mother.
6. The wedding of the congregation member was photographed/catered despite bad weather.
7. The announcement by the factory manager was praised/played by the evening news.
8. The album by the classical composer was praised/played by the radio station.
9. The charge against the Japanese gangster was investigated/withdrawn by the attorney.
10. The accusation of the police officer was investigated/withdrawn by the police chief.
11. The song by the folk singer was heard/enjoyed by a huge crowd.
12. The performance by the Shakespearean actor was heard/enjoyed by all who attended.
13. The verdict for the accused killer was seen/issued by a federal judge.
14. The warrant for the bank robber was seen/issued by the detective.
15. The meeting for the college republican was interrupted/planned by the faculty.
16. The protest for the political activist was interrupted/planned by rival protesters.
17. The statement by the militant terrorist was identified/reprinted by an expert.
18. The script by the screen writer was identified/reprinted by a film critic.
19. The contract for the business executive was threatened/signed during the merger.

20. The agreement for the corporate lawyer was threatened/signed by some investors.
21. The speech by the social worker was applauded/transcribed by the city council.
22. The manuscript by the nuclear physicist was applauded/transcribed by his students.
23. The job for the heart specialist was envied/created by the chief of staff.
24. The position for the teaching applicant was envied/created by the department.
25. The check from the wealthy ambassador was admired/deposited by the treasurer.
26. The donation for the Vietnam veteran was admired/deposited by the veteran's group.
27. The plan of the untraditional architect was criticized/implemented by many engineers.
28. The policy of the local politician was criticized/implemented by the zoning board.
29. The rumor about the school teacher was overheard/spread during the meeting.
30. The secret about the movie star was overheard/spread at the awards ceremony.
31. The recital for the jazz musician was ignored/attended by many fans.
32. The benefit for the AIDS victim was ignored/attended by the press.
33. The curse of the ancient mummy was feared/written by the pharaoh.
34. The prediction of the renown psychic was feared/written by superstitious people.
35. The book about the Russian astronaut was quoted/borrowed by a local expert.
36. The video about the influential author was quoted/borrowed by the English teacher.
37. The documentary about the brave soldier was awarded/financed by the historian.
38. The commercial about the popular athlete was awarded/financed by the shoe company.
39. The article by the newspaper journalist was disliked/typed by the copy editor.
40. The story about the religious leader was disliked/typed by the Catholic church.

References

Barker, J., Nicol, J. L., & Garrett, M. F. (2001). Semantic factors in the production of number agreement. *Journal of Psycholinguistic Research, 30*, 91–114.

Barlow, M. (1999). Agreement as a discourse phenomenon. *Folia Linguistica*, XXXIII, 187–210.

Bock, J. K. (2002). *Persistent structural priming from comprehension to production*. Paper presented at the Fifteenth Annual CUNY Conference on Human Sentence Processing, New York, NY.

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., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology, 43*, 83–128.

Bock, J. K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*, 177–192.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego, CA: Academic Press.

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review, 99*, 150–171.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45–93.

Bock, J. K., Nicol, J. L., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language, 40*, 330–346.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*, B13–B25.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrect, The Netherlands: Foris.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*, 417–437.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology, 42*, 368–407.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychological experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257–271.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*, 313–349.

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science, 23*, 517–552.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A sudy of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1355–1367.

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language, 36*, 147–164.

Eberhard, K. M. (1999). The accessibility of conceptual number to the processes of subject–verb agreement in English. *Journal of Memory and Language, 41*, 560–578.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7*, 195–225.

Fayol, M., Largy, P., & Lemaire, P. (1994). When cognitive overload enhances subject–verb agreement errors: A study in french written language. *Quarterly Journal of Experimental Psychology, 47*, 437–464.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469–495.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Frazier, L., & Clifton, J. C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The relative contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*, 58–93.

Garrett, M. F. (1980). Levels of processing in language production. In B. Butterworth (Ed.), *Language production, vol. 1: Speech and talk* (pp. 177–220). London: Academic Press.

Harm, M. W., Thornton, R., & MacDonald, M. C. (2000). *A distributed, large scale connectionist model of the interaction of lexical and semantic constraints in syntactic ambiguity resolution*. Poster presented at the Thirteenth Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

Hartsuiker, R. J., Anton-Mendez, I., & van Zee, M. (2001). Object attraction in subject–verb agreement construction. *Journal of Memory and Language, 45*, 546–572.

Haskell, T. R., & MacDonald, M. C. (2002). *Proximity does matter: Evidence for distributional effects in the production of subject–verb agreement*. Paper presented at the Fifteenth Annual CUNY Conference on Human Sentence Processing, New York, NY.

Haskell, T. R., & MacDonald, M. C. (in press). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language*.

Hupet, M., Fayol, M., & Schelstraete, M.-A. (1998). Effects of semantic variables on the subject–verb agreement processes in writing. *British Journal of Psychology, 89*, 59–75.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General, 129*, 340–360.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language, 36*, 569–587.

Pearlmutter, N. J. (2000). Linear versus hierarchical agreement feature processing in comprehension. *Journal of Psycholinguistic Research, 29*, 89–98.

Pearlmutter, N. J., Garnsey, S. M., & Bock, J. K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*, 427–456.

Pickering, M. J., Branigan, H. P., Cleland, A. A., & Stewart, A. J. (2000). Activation of syntactic information during language production. *Journal of Psycholinguistic Research, 29*, 205–216.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 940–961.

Pollard, C., & Sag, I. A. (1988). *An information-based theory of agreement*. CLSI Technical Report No. CLSI-88-132. Stanford, CA: CLSI.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior, 22*, 358–374.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24*, 232–252.

Solomon, E. S., & Pearlmutter, N. J. (2002). *The role of semantic integration in syntactic planning in production*. Paper presented at the Fifteenth Annual CUNY Conference on Human Sentence Processing, New York, NY.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effect of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1521–1543.

Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language, 39*, 392–417.

Thornton, R., Haskell, T. R., & MacDonald, M. C. (2001). *A distributional account of agreement production*. Paper presented at the Seventh Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Saarbrucken, Germany.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33*, 285–318.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject–verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition, 61*, 261–298.

Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language, 40*, 455–478.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound and syntax in language production. *Psychological Bulletin, 128*, 442–472.

Vigliocco, G., & Nicol, J. L. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition, 68*, B13–B29.