Journal of Memory and Language 60 (2009) 308–327

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

On the interpretation of the number attraction effect: Response time evidence

Adrian Staub *

Department of Psychology, University of Massachusetts Amherst, Tobin Hall, Amherst, MA 01003, USA

ARTICLE INFO

Article history:
Received 24 June 2008
Revision received 19 November 2008
Available online 30 December 2008

Keywords:
Sentence processing
Agreement
Language production
Response time

ABSTRACT

Speakers frequently make subject–verb number agreement errors in the presence of a local noun with a different number from the head of the subject phrase. A series of four experiments used a two-choice response time (RT) paradigm to investigate how the latency of correct agreement decisions is modulated by the presence of a number attractor, and to investigate the relative latency of errors and correct agreement decisions. The presence of a number attractor reliably increased correct RT, and the size of this RT effect was consistently larger in conditions that also had larger effects on accuracy. Number attraction errors, however, were similar in RT to correct responses in the same experimental condition. These results are interpreted as supporting a model according to which an intervening number attractor makes the agreement computation process more difficult in general [Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making sense of syntax: Number agreement in sentence production. *Psychological Review 112*, 531–559], with errors arising probabilistically. However, attraction from a non-intervening noun resulted in only mildly inflated correct RT, but dramatically inflated error RT, suggesting that non-intervening attraction errors may reflect confusion about the structure of the subject phrase.

Introduction

Over the past two decades, the question of what factors govern agreement behavior has been a major focus of language production research. In large part, this interest is due to the observation that agreement is inconsistent in use, and to the discovery by Bock and Cutting (1992); Bock and Eberhard (1993); Bock and Miller (1991) that it is relatively easy to elicit subject–verb number agreement errors in the laboratory. Bock and colleagues developed a technique in which a participant hears a subject phrase containing a head noun and a subsequent noun (e.g., *the key to the cabinets*) and is instructed to repeat this phrase and continue it however he or she sees fit. Notably, agreement errors tend to occur when the second noun in this phrase (hereafter, the *local noun*) has a different number from the head noun (e.g., **The key to the cabinets are on the table*). This effect is known as *number attraction*. It has since become clear that these errors occur in a range of languages (e.g., Hartsuiker, Schriefers, Bock, & Kikstra, 2003; Vigliocco, Butterworth, & Semenza, 1995; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996), and that subject–predicate gender agreement (Badecker & Kuminiak, 2007; Vigliocco & Franck, 1999, 2001; Vigliocco & Zilli, 1999), pronoun–antecedent gender agreement (Meyer & Bock, 1999), and pronoun–antecedent number agreement (Bock, Eberhard, & Cutting, 2004; Bock, Nicol, & Cutting, 1999), are also error-prone when an "attracting" element is present. Moreover, the presence of a number attractor can also interfere with agreement processing in comprehension (e.g., Nicol, Forster, & Veres, 1997; Pearlmutter, Garnsey, & Bock, 1999).

The initial studies by Bock and colleagues identified several factors that increase or decrease the frequency of number attraction errors. Since these studies, there has been intensive investigation of a wide range of such factors. One robust finding is that errors are more common when the head noun is singular and the local noun is plural

* Fax: +1 413 545 0996.
E-mail address: astaub@psych.umass.edu

0749-596X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2008.11.002

(hereafter, the SP configuration) than when the head noun is plural and the local noun is singular (the PS configuration), known as the *mismatch asymmetry*. This effect has turned out to be quite reliable in English (e.g., Eberhard, 1997), and to be relatively reliable across languages (e.g., Hartsuiker et al., 2003; Vigliocco et al., 1995), though there does appear to be some cross-linguistic variability (see Franck et al., 2004, for discussion). Bock and Eberhard (1993) found that the potency of a local plural as an attractor does not depend on regular plural morphology (i.e., the *–s* ending on the local noun), and that it is not due to the notional plurality of the local noun: Irregular plurals (e.g., *men*) attract, while grammatically singular but notionally plural collective nouns (e.g., *army*) do not. The finding that morphological regularity does not significantly affect the potency of a plural local noun as an attractor has been replicated by Haskell and MacDonald (2003), in English, and by Vigliocco et al. (1995) in Italian. The finding that the notional plurality of the local noun does not drive the effect has been replicated in Dutch and English experiments by Bock, Eberhard, Cutting, Meyer, and Schriefers (2001). In sum, a local plural is an especially effective attractor, and this is due to its grammatical number, rather than to regular plural morphology or to notional number.

Another common finding, initially reported by Bock and Cutting (1992), is that attraction is substantially reduced when the local noun is within an embedded clause, either a noun complement clause (e.g., *The report that they controlled the fires*) or a relative clause (e.g., *The editor who rejected the books*), compared to when the noun is within a prepositional phrase (e.g., *The editor of the history books*). Nicol (1995) and Solomon and Pearlmutter (2004) have since replicated the relative clause finding in English. Experiments in both French and English by Franck, Vigliocco, and Nicol (2002) suggest that this reduction in number attraction may be due to the fact that a local noun within a clausal modifier is relatively deeply embedded in the syntactic structure of the subject phrase, i.e., is structurally quite distant from the root node of the phrase. Franck et al. manipulated the number of each of three nouns in a preamble, in a configuration in which all three were within the same clause (e.g., *The statue(s) in the garden(s) by the mansion(s)*). In these experiments, the second noun, but not the third, was an effective attractor. It appears that a deeply-embedded local noun is a relatively ineffective attractor, whether or not it is in the same clause as the head noun.

The main goal of the present work is to assess a particular view of how number attraction comes about. According to this view, a local noun with a number feature that mismatches the number of the head noun generally makes the process of deciding whether the verb should be singular or plural more difficult, by decreasing the quality, or clarity, of the number information available to the speaker. It is a consequence of this equivocal representation of subject number that the output of the agreement computation process is relatively variable. Putting this slightly differently, the view under consideration proposes that the presence of a number attractor always affects the process of agreement computation, whether or not an error is actually made, and that errors themselves arise on a purely probabilistic basis. As discussed below, this view is implicit in the recent computational model of Eberhard, Cutting, and Bock (2005).

This explanation of number attraction contrasts with the view that is dominant in the literature. It is generally argued that when a speaker erroneously produces a plural verb after a subject phrase like *The key to the cabinets*, she does so because something has gone awry within the agreement computation process on that particular occasion. Specifically, the number feature of the local noun has interfered with the usual process, so that it is this noun's number feature, rather than the head noun's number feature, that is transmitted to the verb. According to what is perhaps the best-known explanation of this sort, errors arise when the number feature on the local noun "percolates" up the syntactic structure of the subject phrase to the root node of this phrase (Bock & Eberhard, 1993; Eberhard, 1997; Franck et al., 2002; Vigliocco et al., 1995). Once this number feature has been instantiated on the root node, it is visible to the mechanism that assigns verb number. Additionally, it is assumed that on those occasions when percolation takes place, the head noun's number is not visible to the agreement computation process, as it is displaced from the root node by the local noun's number.

One motivating virtue of the alternate perspective to be assessed here is that a similar approach has been useful in modeling variable behavior in a wide range of decision situations and tasks. In interpreting data from lexical decision, perceptual categorization, or recognition memory tasks, it is tempting to regard the actual decision on each trial as indicative of an underlying representational state. For example, it appears natural to assume that when a participant in a lexical decision experiment indicates that a given stimulus is a word, a particular word representation has been activated, and this activation has given rise to a "word" response. But in fact, there are impressive models of lexical decision performance (e.g., Ratcliff, Gomez, & McKoon, 2004) which assume that the participant's response is irreducibly probabilistic: While the strength of lexical activation determines the probability that a "word" response will be issued, for any given level of lexical activation it is still the case that some "word" responses and some "nonword" responses will occur. Thus, the very same stimulus may give rise to a "word" response on one trial, and may give rise to a "nonword" response on another trial, with no cognitively interpretable representations or processes differing between the two trials. Similarly, current models of perceptual categorization (Nosofsky & Palmeri, 1997; Nosofsky & Stanton, 2005; Rouder & Ratcliff, 2004) and old–new recognition memory judgment (Nosofsky & Stanton, 2006) assume that response probabilities are determined by the similarity of test items to stored exemplars or prototypes, but that the very same stimulus may be categorized in different ways on different trials, and this does not reflect a difference between trials in the representation of the stimulus itself. Thus, the perspective on subject–verb number agreement to be investigated here assumes commonalities between decision processes in these other cognitive domains and a particular decision process that takes place in the course of language

production. The proposal that such production decisions may have a probabilistic component is not novel, however; Dell (1986), for example, models speech errors such as slips of the tongue as arising from a noisy (i.e., stochastic) spreading activation process.

The remainder of this introduction proceeds as follows. The next section describes the Eberhard et al. (2005) model, as an example of a model claiming that a number attractor exerts a consistent influence on the valuation of subject phrase number, and that the final process of assigning verb number is probabilistic. The subsequent section lays out the response time (RT) predictions that emerge from such a model, and describes previous research bearing on these predictions. Finally, the experiments to be presented here will be previewed.

The marking and morphing model (Eberhard et al., 2005)

The marking and morphing model predicts the probability that a plural verb will be produced on the basis of head noun number, local noun number, and notional factors. The gist of the model is as follows. Number values associated with individual words and syntactic nodes are represented on a scale from 1 (unambiguously plural) to −1 (unambiguously singular). The root node of the subject phrase is marked with its own number value, directly from the message level of the production system, based on the notional number of the intended referent. This value is represented as $S(n)$ in the equation below. In addition, each morpheme in the subject phrase is potentially a source of number information, and this information, when present, is symbolized by the $S(m_j)$ in the equation below. Finally, each morpheme has a weighted connection to the root node, and these are symbolized by w_j. The all-in (or "reconciled") number valuation of the subject phrase, $S(r)$, is the sum of the message-level valuation and the weighted sum of the morpheme valuations.

$$S(r) = S(n) + \sum\nolimits_j (w_j \times S(m)_j)$$

When $S(r)$ is very small or very large, the subject phrase is, for the purposes of computing verb agreement, unequivocally singular or plural, respectively. But when $S(r)$ has an intermediate value, the number of the subject phrase is not clearly singular or plural, but falls somewhere along a singular-to-plural continuum. Number attraction emerges from this model because the value of the summation term depends on number information from both the head noun and any local noun(s). The mismatch asymmetry and structural depth effects emerge as follows. Singular and plural values for count nouns are 0 and 1, respectively, reflecting the assumption (also made by the percolation account) that plural number is grammatically and morphologically marked while singular is an unmarked or default value. Consequently, only plural count nouns affect the summation term. The reduction of attraction when the local noun is deep within the subject phrase follows from an assumption that syntactic connections with more intermediate links have a decreased weight.

It is clear that there are notional number effects on agreement, most notably the tendency to produce plural verbs with collective singular head nouns (e.g., *army*, *team*; Bock et al., 1999, 2004; Humphreys & Bock, 2005), which will be discussed in more detail below. In the Eberhard et al. model, these effects are due to message-level effects on the $S(n)$ term. The model also predicts that head noun collectivity and local noun number should have additive effects on the "all-in" number valuation of the subject phrase, which is generally consistent with the literature (e.g., Haskell & MacDonald, 2003, though cf. Bock et al., 1999). (The potential interaction between these factors is also investigated in the current study; it is important to note that because of the logistic transformation relating $S(r)$ to the probability of a plural verb (described below), additivity at the level of $S(r)$ does not necessarily imply additivity in terms of response proportions.) Furthermore, the model correctly predicts that local noun collectivity should have no effect whatsoever on agreement, due to the fact that local noun collectivity does not affect the $S(n)$ term.

In the Eberhard et al. model, the link between $S(r)$ and verb number is effected by way of a logistic transformation that directly predicts the probability of plural agreement based on the value of $S(r)$: $1/\{1 + \exp - [S(r) + b]\}$, where b is a bias term. Eberhard et al. estimated the values of four free parameters based on data aggregated from many studies of number agreement in English: the value of $S(n)$ when there was notional ambiguity in the subject phrase, the weight parameters for the head noun and the local noun, and the bias parameter. The value of the bias parameter was estimated to be 3.42; thus, any value of $S(r)$ below 3.42 leads to a majority of singular responses, and any value above 3.42 leads to a majority of plural responses. These values were then used to predict response proportions in a series of new experiments, and the model did make impressively accurate predictions relating features of the head noun and the local noun (not all of which have been discussed here) to response proportions.

For present purposes, a critical claim is that only $S(r)$ is accessible to the mechanism that computes verb number; the various values that have entered into the computation of $S(r)$ are invisible to this mechanism. Four aspects of this claim are of special interest. First, the value of $S(r)$ does not vary from one utterance of a given subject phrase to another. Second, $S(r)$ is a continuous rather than discrete variable, and it can take values that are not easily interpreted as either singular or plural. When $S(r) = 4$, for example, we might say that for the purpose of computing agreement, the subject phrase is slightly plural. Third, whether a number attraction error (or, for that matter, an instance of notional agreement with a collective) occurs on a given trial depends only probabilistically, not deterministically, on $S(r)$. Finally, the logistic transformation that relates $S(r)$ to the probability of producing a singular or plural verb is, of course, not intended as a description of *how* $S(r)$ is translated into verb number in a given instance. The equation is a stand-in for a model of a probabilistic process taking $S(r)$ as input and delivering verb number as output.

Response time predictions

The experiments to be presented here were designed to test a set of predictions emerging from the claim that a

number attractor affects the representation of subject phrase number on all trials, with number attraction errors arising from a probabilistic decision process. These predictions relate to the latency of the process of computing verb number. The experiments made use of a paradigm in which it was possible to measure on each trial both the outcome of the agreement computation process and its latency.

The first and most general prediction is that a number attractor should increase the time it takes to compute verb number, even when the speaker ultimately produces the correct verb form. This prediction follows from the assumption common to many decision models (see, e.g., Ratcliff & Smith, 2004) that as the evidence entering into a decision process becomes more equivocal, the process takes longer to complete. In the terms of the marking and morphing model described above, a number attractor makes $S(r)$ closer to the midpoint between singular and plural (i.e., 3.42); when $S(r)$ is close to 3.42, the process of determining verb number will be slowed, however the process ultimately terminates. A second, closely related prediction is that the degree of slowing should correspond to the strength of the number attraction effect: Manipulations that modulate the error rate should similarly modulate the latency of the decision process, with the process being more slowed when the error rate is higher. In the terms of Ratcliff's (1978) well-known diffusion model, for example, manipulations of task difficulty affect the mean rate with which evidence accumulates in favor of the correct response. In fact, in a wide range of cognitive paradigms response latency is found to be even more sensitive than response proportions to manipulations of task difficulty (see Luce, 1986, for extensive discussion).

There is suggestive, if inconclusive, evidence from two previous studies (Haskell & MacDonald, 2003; Nicol et al., 1997) bearing on these issues. Nicol et al. (1997, Experiment 1), using the "maze" buttonpress paradigm (Freedman & Forster, 1985), found that participants were slower to continue a sentence with a correctly number-inflected verb when the subject phrase was in the SP configuration than when the subject phrase was in the SS configuration. However, the form of the effect is puzzling, as RTs in the SP configuration were actually very similar to RTs in the PP and PS configurations.

Haskell and MacDonald (2003) conducted the only study using a spoken production paradigm that has collected both response proportion and RT data. On each trial, the participant read a subject phrase (e.g., *the key to the cabinets*) and an adjective (e.g., *rusty*), and was instructed to produce a question asking whether the subject had the property in question (*Was the key to the cabinets rusty?*). In earlier work, Vigliocco and Nicol (1998) had demonstrated that the rate of attraction errors in the production of such questions is quite similar to the error rate in declarative sentences, even though in the question configuration the local noun does not intervene linearly between the head noun and the verb. Haskell and MacDonald found that a plural local noun increased the proportion of plural auxiliaries, though the attraction effect they obtained was quite small (cf. Vigliocco & Nicol, 1998). Critically, however, a plural local noun also increased speech initiation latency when participants produced a singular auxiliary. (Haskell and MacDonald did not analyze RT for those trials on which a plural was produced, as there were very few relevant data points in some conditions. Though there were 56 participants in their experiment, there were only five observations in each experimental condition, per participant.)

In addition, Haskell and MacDonald manipulated whether the head noun was a singular count noun referring to an individual (e.g., *player*), or a singular collective noun (e.g., *team*). Replicating earlier work, Haskell and MacDonald found that a collective head noun increased the proportion of plural verbs. They also found that speech initiation latency was increased when participants used a singular verb with a collective noun. There was no interaction on the RT measure between head noun collectivity and local noun number; an apparent interaction on the response proportion measure is difficult to interpret, due to ceiling effects, as essentially no errors were made in the easiest condition. The relationship between response proportions and RT was also obtained in a comparison of individual items, with longer RTs occurring for those items that were most likely to elicit variable responses.

Haskell and MacDonald attributed these effects to "partial activation of both singular and plural verb forms" (p. 765), with competition between the two partially-activated forms taking some time to resolve. In fact, there is nothing in these results that directly suggests partial activation of two verb forms, though they do suggest that both the notional number of the head of the subject phrase and the grammatical number of a local noun influence the difficulty of some stage(s) of the process of selecting an appropriate verb form, not only on those trials on which the output of the process clearly reflects the influence of these factors, but also on those trials on which this is not the case.

It is important to note, however, that there is a plausible alternate account of each of the key latency findings, stemming from the fact that sentence initiation latencies in Haskell and MacDonald's experiment may not reflect only the time required to select the form of the fronted auxiliary. These latencies may reflect, in whole or in part, the time required to plan later constituents, specifically the subject phrase itself. If the subject phrase contains mismatching number features, a delay in speech onset could result from the necessity of keeping track of these features. When a participant must remember a subject phrase such as *the actor in the weekend performances*, and then repeat this phrase as part of a question, he or she must encode and retrieve information about which of two distinct number features goes with which noun, and the cost of such memory and retrieval processes cannot be ruled out as an explanation of the effect of a mismatching local noun on speech initiation latency. On the other hand, if the head of the subject phrase is a collective noun such as *cast* or *family*, a delay in speech onset could reflect the time needed to decide whether to construe this noun as distributive, referring to multiple individuals, or as genuinely collective, referring to a group (Humphreys & Bock, 2005). Thus, the delay due to a collective head noun could be attributed to message-level planning, rather than to agreement computation. In sum, Haskell and

MacDonald's latency findings are suggestive, but they require replication in a paradigm in which the latency of verb form selection is not confounded with the latency of subject phrase planning.

Thus, there is already some evidence from previous studies supporting the idea that correct verb agreement decisions are slowed in the presence of a number attractor. However, in one case (Nicol et al., 1997) the data pattern is not entirely clear, while in the other case (Haskell & MacDonald, 2003) the effects appear to be open to alternate interpretations. There is, to date, no evidence bearing on the question of whether the size of any RT effect is modulated by the same factors that increase or decrease the rate of number attraction errors.

An additional prediction that the present experiments were designed to test is that instances in which speakers do and do not make number attraction errors should be associated with essentially similar decision latencies. Error responses in two-choice RT tasks are, in fact, sometimes faster than correct responses, and sometimes slower. Ratcliff and McKoon (2008) point out that until relatively recently modelers had been unable to account for this variability. But Ratcliff, Van Zandt, and McKoon (1999) have shown that the diffusion model can generate both fast errors and slow errors if it is assumed that there is random trial-to-trial variability in the rate of evidence accumulation and in the starting point (i.e., bias) of the decision process. Nevertheless, the most straightforward prediction of an account on which both errors and correct responses emerge from a single probabilistic decision process is that the RTs associated with each response type should be roughly similar. In cognitive terms, a pattern in which error responses are systematically faster than correct responses might suggest that a "fast-guess" process, distinct from the normal process of computing subject–verb number agreement, is operative on some percentage of trials. On the other hand, if error RT is dramatically slower than correct RT, this might suggest that a number attractor occasionally triggers substantial confusion, either about the identity of the head noun or about its number feature, and that errors arise disproportionately from these instances.

In sum, the experiments to be presented here were designed to test three predictions arising from an account of number attraction like that of Eberhard et al., which claims that a number attractor generally affects the difficulty of agreement computation, with errors arising probabilistically: (a) that the presence of a number attractor increases RT for correct responses as well as inducing attraction errors; (b) that the sizes of these two effects are related, with a higher error rate corresponding to more slowing; and (c) that correct and erroneous agreement decisions in the presence of a number attractor are similar in latency. In the general discussion, the question is raised of whether there are other accounts of number attraction, such as the retrieval-based account proposed by Badecker and Kuminiak (2007), that could account for these RT patterns.

Overview of the experiments

To test the predictions just described, four experiments were carried out using a new experimental paradigm. In this paradigm, participants read a subject phrase presented one word at a time at a constant rate (i.e., rapid serial visual presentation, or RSVP; Potter, 1984), then made a speeded decision as to which of two verb forms would provide a grammatical continuation of the sentence. The dependent measures were the response proportions and RT associated with this choice of verb form. A relatively large number of trials were presented in each experimental condition, which allowed for an examination of the relative latency of correct and error RTs, as most participants made number attraction errors.

Experiment 1 replicated Haskell and MacDonald's (2003) design, manipulating head noun collectivity and local noun number. Experiments 2 and 3 investigated whether and how the mismatch asymmetry and syntactic depth effects are manifested in the RT measure. Experiment 4 investigated whether interference from an attractor that is outside the verb's clause (e.g., Bock & Miller, 1991; Franck, Lassi, Frauenfelder, & Rizzi, 2006) has a similar RT profile to the standard type of intervening attraction, addressing the issue of whether non-intervening attraction arises from the same process as standard "intervening" attraction. To preview a critical aspect of the results, the present experiments succeeded, without exception, in replicating the effects on accuracy that have been previously reported in the literature. As will be discussed below, this finding is critical in justifying inferences about the production process based on RT patterns in the present paradigm.

Experiment 1

Replicating the design used by Haskell and MacDonald (2003), two factors were manipulated in Experiment 1: the number of a local noun within a prepositional phrase following the head noun, and whether the head noun denoted an individual or a collective (e.g., *leader* vs. *gang*). A sample set of experimental materials is shown in (1):

(1) a. The leader with the dangerous rival (SS individual)
b. The leader with the dangerous rivals (SP individual)
c. The gang with the dangerous rival (SS collective)
d. The gang with the dangerous rivals (SP collective)

A replication of Haskell and MacDonald's latency effects using the present paradigm would suggest that these effects are not due to the process of planning the subject phrase, and would reinforce the interpretation of these effects as reflecting difficulty related to the process of agreement computation itself. The present experiment included 18 observations per participant in each experimental condition (72 experimental trials overall), intermixed with a large number of fillers, permitting examination of the relative latencies of singular and plural RTs.

Method

Participants

Experiment 1 had 32 participants. In this and all subsequent experiments, participants were members of the University of Massachusetts community who were native

speakers of English, and who were naïve to the experimental hypotheses. Participants in this and all subsequent experiments received either course credit or $10. No one participated in more than one experiment. One participant was excluded from Experiment 1 and replaced due to a large number of responses that exceeded the response deadline.

Materials

Seventy-two experimental fragments were created by adapting items used in three published studies of number agreement in production (Bock et al., 1999; Haskell & MacDonald, 2003; Humphreys & Bock, 2005). Each of these fragments began with the definite article and a noun, followed by a prepositional phrase that contained a local noun, with four versions of each fragment created as described above. The fragments ranged from five to eight words in length, with 6-, 7-, and 8-word fragments including modifiers before either the head noun or the local noun, or both (e.g., *The ballet dancer under the bright lights*). The items were separated into four lists, with each list containing 18 items in each of the four experimental conditions, and with each item appearing in one of its versions on each list. The items used in this and subsequent experiments are presented in the Supplementary materials.

The 72 experimental fragments were combined with 192 filler fragments, also ranging in length from five to eight words. These fillers included 40 fragments with a coordinated subject phrase (e.g., *The sail and the mast*), 40 in which the initial noun was modified by an object-extracted relative clause, while the second noun was the subject of this relative clause (e.g., *The rose bush that the gardeners*), and 40 in which the initial noun was the matrix subject, while the second noun was the subject of a complement clause (e.g., *The diplomat believes that the guards*). For all 120 of these fragments, the correct response was plural. An additional 36 fillers used a structure in which the initial noun was modified by a subject-extracted relative clause (e.g., *The judge who presided over the trial*), and a final 36 used the same structure as the experimental items (i.e., prepositional phrase after head noun), but were designed to investigate an unrelated question that will not be discussed here. For these 72 fillers, the correct response was singular. In sum, the correct response was singular for 144 of the 264 items overall (assuming that singular is regarded as the "correct" verb number with collective head nouns; see below). The 264 items were presented to each participant in an individually randomized order.

Procedure

The experiments were carried out on an IBM-compatible computer running the E-Prime experimental software (Schneider, Eschman, & Zuccolotto, 2002), in a normally-illuminated room, with the participants sitting a natural distance from the keyboard and monitor. Each trial consisted of the following sequence of events. To initiate the trial, the participant pressed the spacebar. Then a fixation cross was presented at the center of the screen for 1 s. After a 150 ms blank screen, the fragment was presented one word at a time, at the center of the screen, in 14-point Times New Roman font. Each word was displayed for 250 ms, with a 150 ms interstimulus interval (ISI), for a total stimulus onset asynchrony (SOA) of 400 ms. After the final 150 ms ISI, the words WAS and WERE were presented on either side of the center of the screen, in uppercase letters. Participants were instructed to select the verb form that would provide a grammatical continuation of the sentence, by pressing either the F or J key on the computer keyboard. For uniformity across participants, the dominant hand was always used for the singular response. If the participant did not respond within 1200 ms of the onset of the response options, a tone was heard and the words "TOO SLOW" appeared on the screen. Pilot work indicated that participants would have little difficulty making a response before this deadline on most trials, and that response accuracy would be within a targeted range, i.e., below ceiling but well above chance. If the participant did make a response within 1200 ms, this response was followed by the word "CORRECT" or "INCORRECT" in the center of the screen.

It is important to note that in order to provide feedback on every trial, it was necessary to code a singular verb response in all four of the experimental conditions as correct, and to code a plural verb response as incorrect. Arguably, plural agreement with collective heads is not incorrect, even for speakers of American English (see Bock et al., 2006), and therefore this feedback might have distorted participants' natural response patterns. However, the rate of plural agreement with collective heads in this experiment was, as noted below, entirely consistent with previously reported estimates in the literature, suggesting that the negative feedback for plural responses did not effectively suppress such responses. For consistency, plural responses will also be referred to here as "errors".

Prior to beginning the experiment, participants read instructions describing the task and emphasizing the importance of both speed and accuracy. Each experimental session began with twelve practice trials, followed by an opportunity to ask questions of the experimenter. Each experimental session lasted approximately 45 min.

Results

Fewer than 1.5% of experimental trials were excluded from the experiment due to responses missing the 1200 ms deadline (32 of 2304 trials). The barplot in Fig. 1 shows the mean correct and error RT in each condition, averaging over individual trials, with error bars representing the standard error of the mean. In this and subsequent figures, correct RT is represented by black bars, and error RT by gray bars. The proportion correct in each condition is also provided in the figure, beneath the condition labels. The main data patterns are clear. The experimental manipulations had approximately additive effects on response proportions, and also had approximately additive effects on correct RT. Error responses occurred more frequently, and correct responses took longer, when the head noun was a collective and when the local noun was plural. Error RT was longer than correct RT in the conditions with a singular local noun, but error and correct RT were very similar in the conditions with a plural local noun.

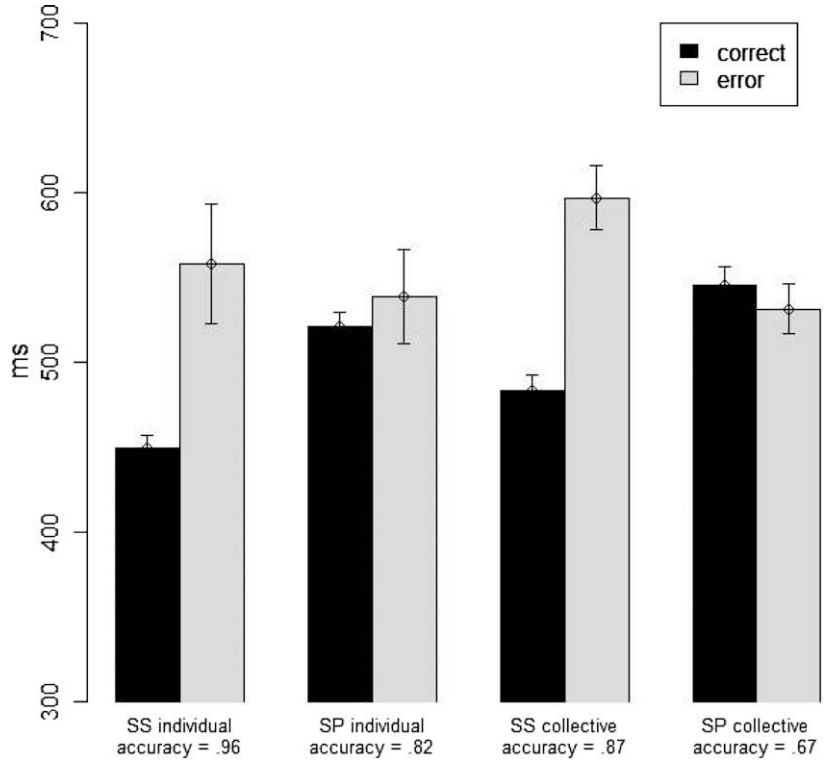

Fig. 1. Experiment 1: mean correct RT and mean error RT in each experimental condition. Error bars represent standard error of the mean. Proportion correct is provided beneath the condition labels.

Statistical analyses of the proportion of correct responses were carried out using mixed-effects logistic regression (Jaeger, 2008), with subjects and items as crossed random factors and with head noun type and local noun number, and their interaction, as fixed effects. This analysis allows simultaneous generalization to populations of subjects and items, and avoids spurious effects that can arise when proportion data are analyzed using traditional ANOVA; see Jaeger (2008) for extensive discussion. Analyses were carried out using *R*, an open-source programming language and environment for statistical computing (R Development Core Team, 2007), and in particular the lme4 package for linear mixed-effects models (Bates, 2005). The results of this analysis are shown in Table 1, which is to be read as follows. The intercept parameter estimate is the log odds of a correct response in the SP collective condition; the fact that this estimate is significant indicates that participants were above chance performance in this condition. The significant parameter estimates associated with an individual head noun, and with a singular local noun, indicate that each of these manipulations further increased the odds of a correct response. Finally, the non-significant, but positive, interaction parameter estimate indicates that though there was a trend toward a further increase in accuracy in the SS individual condition, this increase was not statistically significant.

Table 1
Parameter values for fixed effects in mixed logistic regression model of response proportions in Experiment 1, in log odds, and associated standard errors and probabilities.

	Estimate	Std. error	*p*-Value
Intercept	.875	.188	<.001
Head noun: individual	.897	.149	<.001
Local noun: singular	1.379	.164	<.001
Interaction: individual head noun & singular local noun	.490	.304	.11

To assess experimental effects on the latency of singular responses, a linear mixed model was fit to the data with RT as the dependent measure, with subjects and items as crossed random factors, and with head noun type and local noun number, and their interaction, as fixed effects. Again, this analysis allows simultaneous generalization to populations of subjects and items. The results of the analysis are shown in Table 2; the parameter estimates in this table correspond to millisecond values. An individual head noun decreased correct RT by about 27 ms, and a singular local noun decreased correct RT by about 66 ms; both of these parameter estimates were significant. The estimate of the interaction effect (10 ms) did not approach significance. (Note that all *p*-values were computed based on Markov Chain Monte Carlo sampling, as recommended by Baayen, Davidson, and Bates (2008).)

To compare the latency of correct and error responses, *t*-tests were carried out comparing mean correct RT to mean error RT in each condition, for all subjects who made

Table 2
Parameter values for fixed effects in mixed linear model of singular RT in Experiment 1, in ms, and associated confidence intervals and *p*-values. All confidence intervals and *p*-values obtained by MCMC sampling (Baayen et al., 2008).

	Estimate	95% CI – lower	95% CI – upper	*p*-Value
Intercept	553.59	510.63	594.01	<.001
Head noun: individual	−26.89	−48.80	−3.74	.02
Local noun: singular	−66.04	−88.04	−43.25	<.001
Interaction: individual head noun & singular local noun	−9.81	−40.18	20.39	.53

errors in that condition. These statistical analyses confirmed the patterns in the grand means shown in Fig. 1. Error responses were quite slow in the conditions with a singular local noun. In the SS individual condition, those subjects who made errors averaged 477 ms for correct responses and 566 ms for errors, though due to the small number of errors (and resulting variability) this difference was not significant ($t(13) = 1.7$, $p = .11$). In the SS collective condition, errors were significantly slower than correct responses (564 ms vs. 480 ms; $t(27) = 2.99$, $p < .01$; halfwidth of the 95% CI for the 84 ms difference was 58 ms; hereafter, halfwidths are referred to simply as CIs). In the conditions with a plural local noun, the difference between the latencies of errors and correct responses did not approach significance (SP individual, 545 ms vs. 537 ms: $t(26) < 1$; SP collective, 545 ms vs. 550 ms: $t(29) < 1$).

Discussion

This experiment provided straightforward replication of the response proportion and singular RT effects obtained by Haskell and MacDonald (2003). As in Haskell and MacDonald's experiment, a collective head noun and a plural local noun each increased the frequency of plural responses, with the latter effect being somewhat larger. In both studies, the interpretation of any interaction in the response proportion data is complicated by the potential presence of ceiling effects.

One difference between the results of the present study and Haskell and MacDonald's results is in the overall frequency of plural responses. Haskell and MacDonald reported essentially no errors in the SS individual condition, and only about 1% errors in the SP individual condition. This very low rate of number attraction errors is somewhat surprising, considering that Vigliocco and Nicol (1998), who also had participants produce questions with a fronted auxiliary, found a much higher error rate, similar to the rate of attraction errors in declarative sentences. In the present study, participants made errors almost 4% of the time in the SS individual condition and over 18% of the time in the SP individual condition. Even in the SP collective condition, Haskell and MacDonald's participants used the plural only 11% of the time, while in the present study participants selected the plural in this condition almost 33% of the time. The size of the number attraction effect in the present study is only slightly higher than the average of estimates in the literature, and the present results are also in line with estimates of the rate of plural agreement with collective heads (see, e.g., the meta-analysis conducted by Eberhard et al. (2005)).

As in Haskell and MacDonald's study, the present experiment found significant slowing of correct responses when the head noun was a collective, and when the local noun was plural; and as in that study, there was no convincing evidence of an interaction effect on correct RT. The effect of local noun number on correct RT was more than twice the size of the effect of head noun type, which is once again consistent with the results reported by Haskell and MacDonald. (Haskell and MacDonald did not report the effect sizes directly, but based on inspection of their Fig. 1, it appears that the effect of a plural local noun was about twice as large as the effect of a collective head noun.) Thus, the RT effects reported by Haskell and MacDonald appear to be reliable, and appear to generalize to an experimental paradigm with rather different characteristics.

The new finding in this experiment relates to the comparison of correct RT and error RT. Number attraction errors (i.e., errors made in the presence of a plural local noun) were not significantly different in their latency than the corresponding correct responses. On the other hand, error RT was slower than singular RT in the singular local noun conditions. The increased latency of errors has a clear explanation in the SS individual condition, as here it is plausible that errors would tend to occur primarily (or only) when the participant has been distracted, or when her attention has lapsed. Recently, Ratcliff and Tuerlinckx (2002) have explicitly added a "contaminant" parameter to the diffusion model to account for such anomalous responses, which would tend to occur with equal probability in all conditions, but would be likely to make up a large proportion of the error responses when errors are quite rare. In the SS collective condition, contaminant responses alone are unlikely to account for the slowing of errors. In this condition, plural responses may arise from a process that is not, in fact, identical to the process that leads to singular responses. It is possible that on some proportion of trials, the presence of a collective head noun triggers a relatively slow decision process (perhaps a process of deciding whether to conceptualize the referent of the head noun in a collective or distributive manner; Humphreys & Bock, 2005), and that plural responses arise disproportionately from this slow process. From the present perspective, it is important to note that the finding of inflated error RT in the SS conditions suggests that the lack of such a finding in the SP conditions is not due to a lack of sensitivity of the experimental paradigm, or to a lack of statistical power.

Experiments 2 and 3 tested whether factors that modulate the rate of number attraction errors (i.e., the mismatch asymmetry and the syntactic depth effect) also modulate the degree of slowing of correct RT. This result would be expected on the view that number attraction errors and slowing of correct responses are two overt indices of a single underlying factor affecting the difficulty of agreement decisions, such as the value of $S(r)$ in the Eberhard et al.

(2005) model. Experiments 2 and 3 also provided additional tests of the finding that number attraction errors do not differ in their latency from correct responses in the presence of an attractor.

Experiment 2

Experiment 2 tested whether the effect of a number attractor on decision latency is less pronounced when the head noun is plural and the local noun singular (the PS configuration), as in this configuration the rate of attraction errors tends to be reduced, as discussed above. In the critical sentences in Experiment 2, the subject phrase contained both a head noun and a local noun, as in Experiment 1. The head noun was either singular or plural, and the local noun either matched or mismatched the head noun in number. These conditions are illustrated in (2a–d) below:

(2) a. The door to the office (SS)
b. The door to the offices (SP)
c. The doors to the offices (PP)
d. The doors to the office (PS)

Method

Participants

There were 24 participants in Experiment 2.

Materials

One hundred sixty experimental quadruplets like (2a–d) above were constructed. These were adapted from items used in five previous studies of subject–verb number agreement (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991; Bock et al., 2004; Vigliocco & Nicol, 1998). As in Experiment 1, the fragments ranged from five to eight words in length. Four versions of each fragment were constructed, by manipulating head noun and local noun number as described above. The items were separated into four lists, with each list containing 40 items in each of the four experimental conditions, and with each item appearing in one of its versions on each list.

The 160 experimental fragments were combined with 80 filler fragments. These fillers were made up of 40 of each of two types used in Experiment 1: The initial noun could be the head of a relative clause, while the second noun was the subject of this relative clause (e.g., *The rose bush that the gardener*), or the initial noun could be the matrix subject, while the second noun was the subject of a complement clause (e.g., *The diplomat believes that the guard*). These particular filler types were included in order to require participants to attend to the number of the second noun phrase in the fragment, as with these constructions it is the second noun phrase that controls agreement with a subsequent verb. Singular was the correct response for half of the items of each of the two filler types, so that in the experiment as a whole each of the two response options was correct on 50% of trials. The 80 filler items and the 160 experimental items were presented in an individually randomized order to each participant.

Procedure

The procedure was identical to Experiment 1, with the exceptions that: (a) the words IS and ARE were used as response options, instead of WAS and WERE and (b) the location of each of the response options was rotated so that half of the participants used their dominant hand for singular, and half for plural. There were no significant main effects or interactions related to this between-subjects manipulation, so it is left out of all analyses below.

Results

Fewer than 1% of experimental trials were excluded due to responses exceeding the 1200 ms deadline (28 of 3840 trials). One experimental item was excluded altogether, due to a programming error. The barplot in Fig. 2 shows the mean correct and error RT in each condition (with standard errors), and provides the proportion correct responses in each condition.

As for Experiment 1, logistic regression was used to assess the effects of the experimental manipulations on response proportions (Table 3). As expected, there was a significant effect of a mismatching local noun on accuracy ($p < .001$), and also a significant interaction effect ($p < .001$), with a larger mismatch effect for the singular head conditions than for the plural head conditions. As the proportions in Fig. 2 indicate, this interaction was driven in part by lower accuracy in the SP condition than in the PS condition, and in part by higher accuracy in the SS condition than in the PP condition.

Table 4 displays the results of the mixed linear model analysis of correct RT. As for the accuracy data, there was a significant effect of match vs. mismatch on RT, and also a significant interaction effect. As Fig. 2 indicates, mean correct RT was very similar in the SS and PP conditions (435 ms and 436 ms, respectively) but substantially longer in the SP condition than in the PS condition (533 ms and 492 ms, respectively).

As for Experiment 1, error RT was compared to correct RT by carrying out *t*-tests on subject means in each condition. As in Experiment 1, error RT was longer than correct RT in the SS condition (522 ms vs. 438 ms), though again, this sizable difference was not significant due to the small number of errors ($t(12) = 1.82$, $p = .09$). However, error RT was significantly longer than correct RT in the PP condition (604 ms vs. 437 ms; $t(21) = 3.97$, $p < .001$; 95% CI was 88 ms). In both mismatch conditions, error RT was numerically slower than correct RT, though in neither condition did this difference reach full statistical significance (SP: 554 ms vs. 530 ms, $t(23) = 1.59$, $p = .13$, 95% CI was 31 ms; PS: 539 ms vs. 496 ms, $t(23) = 2.05$, $p = .053$, 95% CI was 44 ms).

Discussion

Beginning with the accuracy data, Experiment 2 successfully replicated the basic number attraction and mismatch asymmetry effects: More agreement errors were made in the presence of a mismatching local noun, and this mismatch effect was stronger when the head noun was singular than when the head noun was plural. These find-

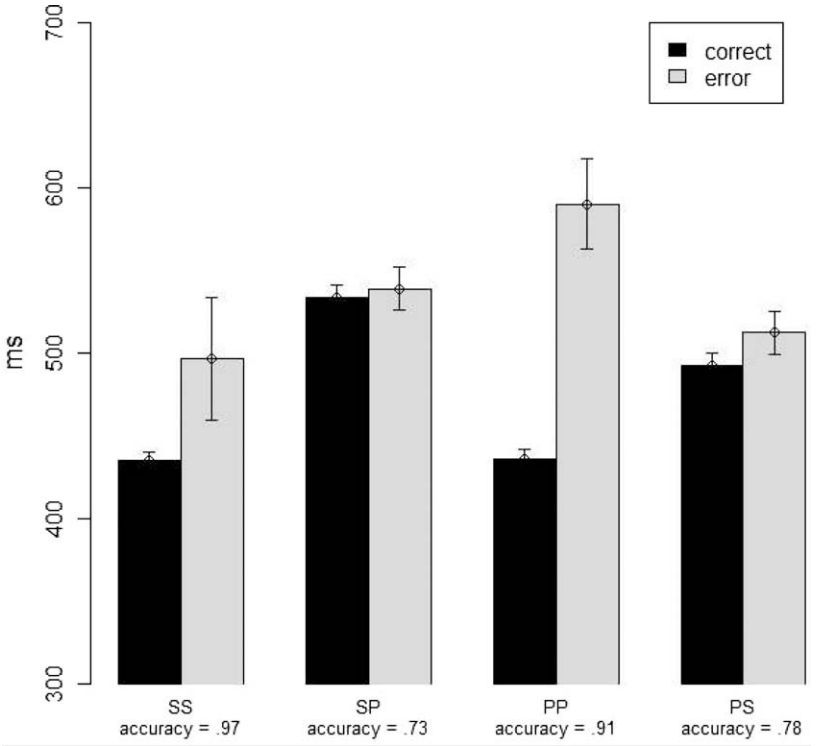

Fig. 2. Experiment 2: mean correct RT and mean error RT in each experimental condition. Error bars represent standard error of the mean. Proportion correct is provided beneath the condition labels.

Table 3
Parameter values for fixed effects in mixed logistic regression model of response proportions in Experiment 2, in log odds, and associated standard errors and probabilities.

	Estimate	Std. error	*p*-Value
Intercept	2.587	.193	<.001
Head noun: singular	1.40	.243	<.001
Local noun: mismatch	−1.11	.144	<.001
Interaction: singular head noun & mismatching local noun	−1.764	.268	<.001

Table 4
Parameter values for fixed effects in mixed linear model of singular RT in Experiment 2, in ms, and associated confidence intervals and *p*-values. All confidence intervals and *p*-values obtained by MCMC sampling (Baayen et al., 2008).

	Estimate	95% CI – lower	95% CI – upper	*p*-Value
Intercept	438.03	394.66	482.10	<.001
Head noun: singular	−2.51	−17.87	12.87	.73
Local noun: mismatch	54.63	38.06	70.59	<.001
Interaction: singular head noun & mismatching local noun	44.58	21.43	68.22	<.001

ings are important in validating the experimental paradigm. The fact that number attraction exists in this paradigm and the fact that its strength depends on head number support the inference that in choosing a grammatical continuation in the present paradigm, participants are computing agreement using mechanisms that are similar to those that are employed in the standard laboratory production paradigm.

Two aspects of the accuracy data are worth noting. First, more errors were made in the PP condition than in the SS condition. Thus, the critical interaction effect was due only in part to a finding of more errors in the SP condition than in the PS condition, and in part to a difference between the "baseline" match conditions. Second, there were number attraction errors in the PS condition as well as in the SP condition. Summing up these two issues, accuracy might have been expected to be closer to ceiling in both the PP and PS conditions.

In fact, these findings are consistent with the literature as a whole. In some English production studies (e.g., Bock & Cutting, 1992; Bock & Miller, 1992; Eberhard, 1997; Franck et al., 2002; Vigliocco & Nicol, 1998) as well as studies in other languages (e.g., Vigliocco et al., 1995, for Italian; Franck et al., 2006, for French) there is at least some difficulty in the PP condition (indeed, in one experiment Franck et al., 2006, somewhat paradoxically found the highest rate of agreement errors in this condition; for discussion of this phenomenon, see Franck et al., 2002,2004). Moreover, it is not the case that the mismatch asymmetry always takes the form of a substantial error rate in the SP condition and no agreement errors in the PS condition. While this is sometimes the pattern, it is also common for researchers to report some errors in the PS condition and there are even isolated studies reporting little difference between the SP and PS conditions (e.g., Franck et al., 2002, for both

French and English). In sum, the finding of an accuracy ordering SS > PP > PS > SP is consistent with the overall picture that emerges from the literature. A critical point is that in the present experiment, the error rate was relatively high overall. This suggests that accurate performance in the paradigm used here may be slightly more difficult than in the standard production paradigm, allowing performance in all conditions to come off ceiling, and allowing the four-way distinction between conditions to become more clearly evident.

Turning to the RT data, Experiment 2 clearly replicated the finding of increased correct RT when the local noun mismatches the head noun in number. Moreover, this RT effect was significantly larger in the singular head conditions than in the plural head conditions; in other words, the magnitude of the RT effect reflected the mismatch asymmetry. Apparently, correct RT closely tracks changes in accuracy across conditions, with lower accuracy corresponding to longer RT.

As in Experiment 1, error RT was much slower than correct RT on the relatively rare occasions on which an error was made in the absence of a number attractor. Error RT was also numerically slower than correct RT in the mismatch conditions, but this difference was not fully significant in either condition. Experiments 3 and 4 provide additional opportunities to assess the relative speed of errors and correct responses in the presence of mismatching local noun.

Finally, the comparison between the SS and PP conditions is also notable. Though errors were more likely in the PP condition, correct RT was essentially identical in the two conditions. And as noted above, when errors were made in the PP condition they tended to be dramatically slower than correct responses. This suggests that errors in the PP condition, unlike errors in the mismatch conditions, may not be due to the general difficulty of the decision process in this condition, but to difficulty that is triggered (for some reason) on a small minority of trials.

Experiment 3

The main purpose of Experiment 3 was to assess whether the reduction in the rate of number attraction errors associated with a deeply-embedded local noun is also associated with a reduced effect on correct RT.

Method

Participants

There were 36 participants in Experiment 3.

Materials

The items in Experiment 3 consisted of modified versions of 144 of the 160 items used in Experiment 2. Four critical experimental conditions were created. In all of these conditions the head noun was singular, while the local noun could be either singular or plural. In two of the conditions, the local noun was within a prepositional phrase that was either a modifier or an argument of the head noun, as in (3a and b) below. Two modifiers preceded the local noun, resulting in a fragment that was seven words in length. In the other two conditions, the local noun was the final word of a subject-extracted relative clause following the head noun, as in (3c and d). These versions were also seven words in length, and the central questions involved the comparison between these length-matched conditions. Two additional conditions (3e and f) were also included, which were identical to conditions (a and b) except that the modifiers preceding the local noun were eliminated, so that the fragments were five words in length. The purpose of these additional conditions was primarily to reduce the predictability of the timing of the onset of the response options, and results from these conditions will not be discussed below. (While comparisons between conditions (a and b) and conditions (e and f) might have been used to explore the issue of how prepositional phrase length affects number attraction (Bock & Cutting, 1992), additional experiments not reported here suggested that in the present paradigm there are sizable, and difficult to interpret, effects of the length of the fragment preceding the response options on participants' readiness to respond, reflected in RT.)

(3) a. The judge for the important criminal trial (SS)
b. The judge for the important criminal trials (SP)
c. The judge who presided over the trial (SS relative clause)
d. The judge who presided over the trials (SP relative clause)
e. The judge for the trialf. The judge for the trials.

The items were separated into six lists, with each list containing 24 items in each of the six conditions.

The 144 experimental fragments were combined with 120 fillers. These fillers included 40 fragments with a coordinated subject phrase (e.g., *The sail and the mast*), 40 in which the initial noun was modified by an object-extracted relative clause, while the second noun was the subject of this relative clause (e.g., *The rose bush that the gardeners*), and 40 in which the initial noun was the matrix subject, while the second noun was the subject of a verbal complement clause (e.g., *The diplomat believes that the guards*). For all 120 of these fragments, the correct response was plural. The 144 experimental items and 120 fillers were presented in an individually randomized order to each participant.

Procedure

The procedure was identical to Experiment 1.

Results

Fewer than 1% of experimental trials were excluded due to responses missing the 1200 ms deadline (36 out of 5184). An additional 16 trials were eliminated due to a programming error. Fig. 3 displays correct and error RT for the critical conditions, along with the proportion correct in each condition.

The results of a logistic regression on the proportion correct responses are shown in Table 5. A mismatching local noun again reduced the probability of a correct response. There was a significant interaction with structure type, as an attractor within a relative clause had a smaller

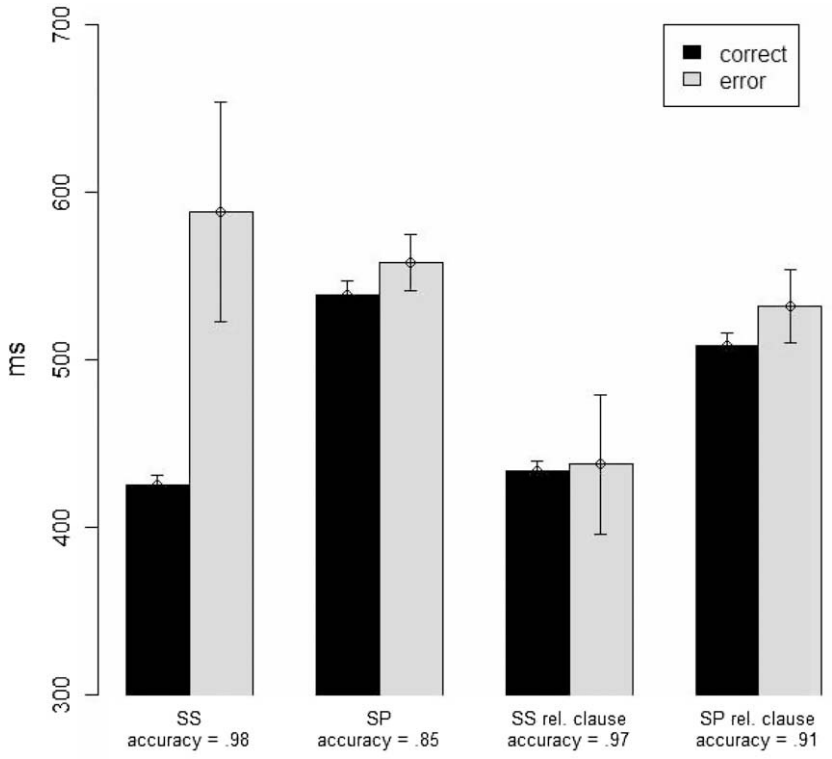

Fig. 3. Experiment 3: mean correct RT and mean error RT in each experimental condition. Error bars represent standard error of the mean. Proportion correct is provided beneath the condition labels.

effect than one within a prepositional phrase. The same pattern is seen in the analysis of correct RT (Table 6). A mismatching local noun increased correct RT, but this increase was smaller when the local noun was within a relative clause than when it was within a prepositional phrase.

Turning to the comparison of error RT and correct RT, errors in the SS condition were again slower than correct responses, though nonsignificantly so (556 ms vs. 454 ms; $t(9) = 1.24$, $p = .25$). Errors were also nonsignificantly slower than correct responses in the SS relative clause condition (465 ms vs. 442 ms; $t(8) < 1$). Errors were marginally faster than correct responses in the SP condition (527 ms vs. 568 ms; $t(30) = 1.95$, $p = .06$, 95% CI was 43 ms), and were nonsignificantly faster than correct responses in the SP relative clause condition (526 ms vs. 538 ms; $t(27) < 1$).

Table 5
Parameter values for fixed effects in mixed logistic regression model of response proportions in Experiment 3, in log odds, and associated standard errors and probabilities.

	Estimate	Std. error	*p*-Value
Intercept	4.401	.303	<.001
Type: relative clause	−.308	.335	.36
Local noun: mismatch	−2.33	.273	<.001
Interaction: relative clause & mismatching local noun	.928	.371	.01

Discussion

The results from Experiment 3 are straightforward. Like the mismatch asymmetry, the modulation of number attraction by syntactic depth was replicated in the current paradigm. In addition, there was again a correspondence between error rate and correct RT. A number attractor in a relative clause induced fewer errors than one within a prepositional phrase, and also had a smaller effect on the latency of correct responses. As in Experiments 1 and 2, number attraction errors were not significantly different in their latency from the corresponding correct responses. Across the six mismatch conditions in the three experiments, errors were numerically slower than correct responses four times, and were numerically faster than correct responses twice.

Table 6
Parameter values for fixed effects in mixed linear model of singular RT in Experiment 3, in ms, and associated confidence intervals and *p*-values. All confidence intervals and *p*-values obtained by MCMC sampling (Baayen et al., 2008).

	Estimate	95% CI – lower	95% CI – upper	*p*-Value
Intercept	426.28	388.70	463.87	<.001
Type: relative clause	8.32	−7.44	23.69	.30
Local noun: mismatch	123.07	106.93	139.16	<.001
Interaction: relative clause & mismatching local noun	−45.77	−68.89	−22.93	<.001

Together, the results from Experiments 2 and 3 argue against certain very simple explanations of the effect of a number attractor on response latency. For example, it might have been proposed, based on Experiment 1 (and on Haskell & MacDonald, 2003) that the presence of a mismatching local noun slows responding simply because, when this noun is present, correct verb agreement requires inhibiting a response to an irrelevant aspect of the stimulus, which increases RT in a wide range of paradigms, such as the Stroop task (Stroop, 1935). But Experiments 2 and 3 illustrate that the strength of the effect of a mismatching local noun on RT is modulated by the mismatch asymmetry and by syntactic depth. Thus, an account claiming that the RT effect is due to inhibition of a response to an irrelevant distractor must explain why it is harder to inhibit a response to a plural local noun than to a singular one, and why it is easier to inhibit a response to a local noun that is within a clausal modifier. The critical variables that determine RT appear to relate to linguistic features and linguistic structure, rather than to superficial stimulus properties.

Finally, it is reasonable to ask how regular is the relationship between response proportions and RT across the 12 conditions in Experiments 1–3. This relationship is illustrated in panel (a) of Fig. 4, which plots overall RT against the probability of a plural response for these 12 conditions. There is indeed a rather regular relationship, with RT increasing as the probability of a plural response moves closer to .5, even though these data were gathered from three distinct groups of subjects, with distinct sets of items, and therefore a substantial amount of between-experiment variability would be expected. However, there is a suggestion of non-linearity in the relationship. Panel (b) of Fig. 4 plots overall RT against $S(r)$, computed from the probability of a plural response by using Eberhard et al.'s (2005) logistic transformation, adopting their value for the scaling parameter of $b = 3.42$. Interestingly, visual inspection suggests that $S(r)$ is a better linear predictor of RT (though there are far too few data points to conduct a convincing test of whether the difference in linear correlation is statistically significant). This pattern would be expected, on the assumption that the relationship between response proportions and RT is indirect, with the value of each of these variables directly reflecting $S(r)$.

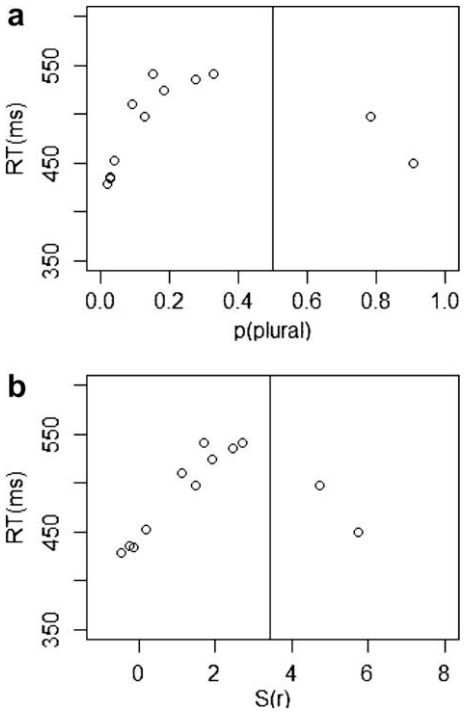

Fig. 4. Response time in the 12 conditions in Experiments 1–3 as a function of the probability of a plural response (panel a) and as a function of $S(r)$ (panel b). The vertical line in each graph corresponds to p (plural) = .5.

Experiment 4

It is clear from the results of both Vigliocco and Nicol (1998) and Haskell and MacDonald (2003) that number attraction can occur in a configuration in which the attractor does not intervene linearly between the head of the subject phrase and the inflected verbal element. The finding that a fronted auxiliary in a yes–no question is subject to number attraction can be explained on the assumption that agreement is computed on an intermediate representation constructed in the course of grammatical encoding, in which constituents have not yet been assigned their final linear ordering (Vigliocco & Nicol, 1998). What is more surprising is the existence of number attraction in a configuration in which the attracting noun intervenes neither in the surface representation nor in any posited intermediate representation. Such apparently non-intervening attraction was investigated in the very first experimental study of number attraction (Bock & Miller, 1991, Experiment 3; cf. Kimball & Aissen, 1971). Bock and Miller presented preambles that consisted of a noun followed by a relative pronoun and a relative clause subject (e.g. *The colonies that the king*), so that a completion required both a relative clause verb (with *the king* as subject) and a main clause verb (with *the colonies* as subject). In contrast to yes–no questions, on the standard analysis of such object-extracted relative clauses there is no underlying "pre-movement" representation in which the initial noun intervenes between the relative clause subject and the relative clause verb (though see the discussion of Franck et al., 2006, below). Participants made frequent number agreement errors on the relative clause verb when the initial noun and relative clause subject mismatched in number. As usual, this effect was strongest when the attracting noun (in this case, the initial noun) was plural.

Interestingly, Bock and Miller also found that the error rate in this configuration was affected by the animacy of the initial noun, with many more errors appearing when this noun was animate, and therefore a plausible agent. In previous experiments, they had found that animacy did not have a similar effect on the error rate in the standard configuration in which the attracting noun appeared in a prepositional phrase postmodifier (a conclusion that has

since been confirmed by Barker, Nicol, & Garrett, 2001). Bock and Miller suggested that in the relative clause configuration speakers may suffer from a "subject identification problem," i.e., on some proportion of trials they are genuinely confused about which of the two nouns in the preamble should be the subject of the first verb produced. An examination of the content of the elicited completions provided additional support for this view, as in some cases participants produced completions that implied that they were treating the first noun as the subject of the first verb. Moreover, in these cases an agreement error was especially likely.

There is some independent plausibility to the idea that speakers may have difficulty keeping track of incomplete syntactic dependencies in the configuration tested by Bock and Miller. The difficulty associated with comprehension of object-extracted relative clauses is well known (e.g., King & Just, 1991; Traxler, Morris, & Seely, 2002), and has often been attributed to working memory constraints (e.g., Gibson, 1998; Waters & Caplan, 1996). According to such memory-based accounts, an object relative (e.g., *The colonies that the king ruled asked for their independence*) may be harder to understand than a corresponding subject relative (e.g., *The colonies that disobeyed the king asked for their independence*) because in the former case, two noun phrases must be maintained simultaneously in working memory, tagged to their respective syntactic positions, while the processor awaits the corresponding verb phrases. In support of the idea that memory load and memory interference play a role in the comprehension of object relatives, Gordon, Hendrick, and Johnson (2001), Gordon, Hendrick, and Johnson (2004) have found that comprehension difficulties are greatest for object relatives when the two noun phrases are of the same type, i.e., both names or both definite descriptions. Furthermore, in a parallel to Bock and Miller's (1991) production results, Mak, Vonk, and Schriefers (2002), Mak, Vonk, and Schriefers (2006); see also Gennari & MacDonald, 2008; Traxler, Williams, Blozis, & Morris, 2005 have found that the difficulty associated with object relatives is especially pronounced when the initial noun is animate (e.g., *The director that the movie pleased*). Thus, Bock and Miller's findings fit well with the comprehension literature, on the assumption that working memory limitations and animacy may play a role in the production process as well.

More recently, Franck et al. (2006, Experiment 3) reported similar attraction in a French experiment, in which speakers were to provide relative clause verbs after cleft preambles such as *C'est les negociations que le ministre* ("It's the negotiations that the minister"). However, Franck et al. assimilated this type of attraction to the usual intervening variety, arguing that there is in fact an intermediate stage in the syntactic derivation in which the object noun phrase (*les negociations*) intervenes between the relative clause subject (*le ministre*) and the verb (Chomsky, 1995; Kayne, 1989). Unlike Bock and Miller, Franck et al. did not find a mismatch asymmetry, though in fact this asymmetry has often proved to be weak or nonexistent in French (e.g., Franck et al., 2002; Franck et al., 2004).

The present experiment was designed to arbitrate between Bock and Miller's (1991) original speculation about the source of attraction in the object relative clause configuration, and the account more recently offered by Franck et al. (2006). On the Franck et al. account, there is no reason to assume that the pattern of effects of a non-intervening attractor on accuracy and RT should be qualitatively different from the pattern of effects caused by an intervening attractor, as it is posited by Franck et al. that the attractor in the relative clause configuration does in fact intervene, at the stage at which agreement is computed. On Bock and Miller's account, on the other hand, agreement errors arise in the object-extracted relative clause configuration because, due to the complexity of the construction, the speaker is sometimes confused about which noun phrase is in fact the subject of the relative clause verb. Thus, the Bock and Miller account allows for distinct data patterns to emerge; specifically, it suggests that errors due to a non-intervening attractor might differ in RT from the corresponding correct responses, on the assumption that error trials involve distinct processing problems. On the other hand, the Bock and Miller account suggests that correct responses in the presence of a non-intervening attractor might show little slowing compared to correct responses in the corresponding match condition, as there is no mismatching local noun within the subject phrase capable of exerting a regular influence on the all-in valuation of subject phrase number.

Method

Participants

There were 24 participants in Experiment 4. Three participants were excluded and replaced, one due to performance that was not significantly above chance, and two because in debriefing they reported adopting specific strategies based on key words.

Materials

The items in Experiment 4 were created by adapting the 160 items used in Experiment 2. The SS and SP conditions were retained from that experiment, and two new conditions were created by inverting the order of the head noun and local noun, making the local noun into the initial noun and the head noun into the relative clause subject. The four conditions are shown in (4a–d) below. Note that condition (d) is referred to as the SP non-intervening condition, despite the fact that the plural noun linearly precedes the singular noun. In what follows, this plural noun will also be referred to as the "local" noun, for ease of exposition.

(4) a. The key to the cabinet (SS)
b. The key to the cabinets (SP)
c. The cabinet that the key (SS non-intervening)
d. The cabinets that the key (SP non-intervening)

In order to construct semantically sensible relative clause conditions, 19 of the 160 items had to be modified, usually by changing one of the two nouns.

Four counterbalancing lists were created with 40 items in each condition on each list. The 160 experimental items

were combined with 80 filler items, for which plural was the correct response. Twenty of these fillers used a prepositional phrase structure similar to that in (4a and b), and twenty used a relative clause structure similar to that in (4c and d). An additional twenty used a coordinated subject phrase, and a final twenty used a sentence complement structure. The experimental items and fillers were presented in an individually randomized order to each participant.

Procedure

The procedure was identical to Experiments 1 and 3.

Results

About 1% of experimental trials were excluded due to responses missing the 1200 ms deadline (39 out of 3840). Fig. 5 displays the mean correct and error RT in each condition, as well as the proportion correct responses. Several departures from previous patterns are evident. First, while the effect of a number attractor on accuracy was roughly similar in the intervening and non-intervening configurations, the effect on correct RT was more than twice as large in the intervening configuration as in the non-intervening configuration. This is due to much longer RT in the SS non-intervening condition than in the SS condition, but more similar RT in the two SP conditions. Second, error RT is again very similar to correct RT in the SP condition, but is clearly longer than correct RT in the SP non-intervening condition.

Mixed-effects logistic regression on the proportion correct responses (Table 7) confirms that the size of the attraction effect on accuracy was not significantly smaller in the non-intervening conditions than in the intervening conditions; only the main effect of local noun number was significant. However, a mixed linear model with correct RT as the dependent measure (Table 8) reveals a highly significant interaction. The non-intervening configuration and the presence of a number attractor each increased RT, but these effects were underadditive, with the number attraction effect being much smaller in the non-intervening conditions.

Turning to the comparison of the latencies of errors and correct responses, error RT was again very slow in the SS condition (547 ms vs. 431 ms), and on this occasion the difference was significant ($t(10) = 2.39$, $p < .05$, 95% CI was 108 ms). Error RT was also slow compared to correct RT

Table 7
Parameter values for fixed effects in mixed logistic regression model of response proportions in Experiment 4, in log odds, and associated standard errors and probabilities.

	Estimate	Std. error	*p*-Value
Intercept	3.944	.263	<.001
Type: non-intervening	−0.278	.281	.32
Local noun: mismatch	−2.374	.227	<.001
Interaction: non-intervening & mismatching local noun	.520	.307	.09

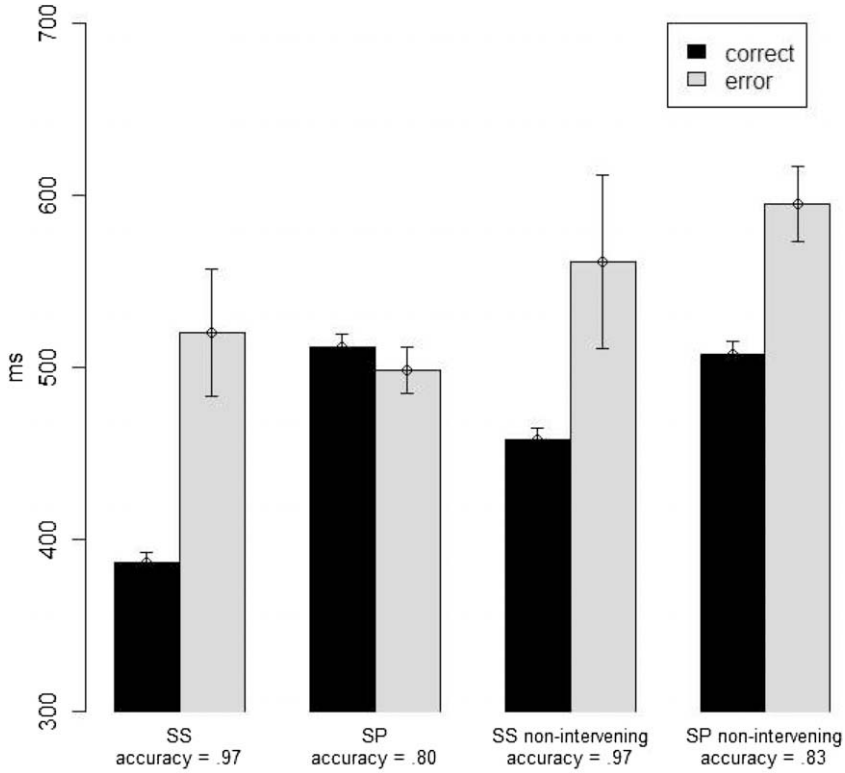

Fig. 5. Experiment 4: mean correct RT and mean error RT in each experimental condition. Error bars represent standard error of the mean. Proportion correct is provided beneath the condition labels.

Table 8
Parameter values for fixed effects in mixed linear model of singular RT in Experiment 4, in ms, and associated confidence intervals and *p*-values. All confidence intervals and *p*-values obtained by MCMC sampling (Baayen et al., 2008).

	Estimate	95% CI – lower	95% CI – upper	*p*-Value
Intercept	388.44	345.86	433.60	<.001
Type: non-intervening	69.49	54.02	84.70	<.001
Local noun: mismatch	122.87	106.10	139.11	<.001
Interaction: non-intervening & mismatching local noun	−72.04	−93.35	−47.26	<.001

in the SS non-intervening condition (569 ms vs. 450 ms; $t(14) = 3.21$, $p < .01$, 95% CI was 80 ms). Again, error RT was not significantly different from correct RT in the SP condition (501 ms vs. 516 ms; $t(22) < 1$). However, error RT was significantly slower than correct RT in the SP non-intervening condition (589 ms vs. 511 ms; $t(22) = 2.53$, $p < .02$, 95% CI was 64 ms).

Discussion

The results from this experiment may be summarized as follows. The error rate in the presence of a non-intervening attractor was not significantly different from the error rate in the presence of an intervening attractor. However, the RT patterns were very different. First, the slowing of correct responses in the SP non-intervening condition, compared to the SS non-intervening condition, was much less pronounced than the slowing in the SP condition, compared to the SS condition. Thus, there was a highly significant interaction effect on correct RT. Second, errors made in the presence of an intervening attractor were not slower than the corresponding correct responses (indeed, they were numerically faster), but errors made in the presence of a non-intervening attractor were significantly slower than the corresponding correct responses.

These results support an account of non-intervening attraction according to which this type of attraction is not assimilated to the usual (i.e., intervening) variety. While non-intervening attraction looks like intervening attraction in terms of the rate of agreement errors it induces, it is clearly distinct from intervening attraction in terms of its effect on processing, when RT is taken into account. A non-intervening attractor has a relatively small effect on the duration of the agreement computation process on those trials on which the correct verb form is ultimately chosen, while having a rather pronounced effect on the duration of the agreement computation process on those trials on which the incorrect form is chosen. One explanation of this pattern would be by way of Bock and Miller's (1991) idea that when a speaker must produce the relative clause verb in an object relative construction, he or she occasionally experiences a "subject identification problem." On a minority of trials, the speaker loses track of which noun phrase is the matrix subject and which is the relative clause subject, and resolving this confusion (a process which is often unsuccessful) takes considerable time. The cases in which this subject identification problem occurs are disproportionately cases in which an error is ultimately made.

One caveat regarding the interpretation of this experiment is that it is possible that comprehension difficulties played some role in generating the pattern of results. As noted above, object relative clauses are notoriously difficult for comprehenders. Indeed, it is clear that the non-intervening configuration increased RT, even in the SS condition. Furthermore, subject identification problems would likely be more prevalent in comprehension than in production, where the speaker presumably knows, at least at the outset, which noun phrase was intended as the matrix subject, and which as the relative clause subject. Thus, it seems especially important to replicate the RT results of Experiment 4 in a more traditional production paradigm.

General discussion

The principal results from the preceding four experiments were as follows. In all experiments, the presence of an intervening number attractor reliably induced errors when participants were required to choose a number-inflected verb form that would provide a grammatical continuation of a subject phrase presented in RSVP format. In addition, a number attractor reliably increased the latency of correct responses. The error rate was modulated by factors that are known to affect the error rate in spoken production studies (i.e., the mismatch asymmetry, the syntactic depth effect). The effect of a number attractor on correct RT was also modulated by these factors, as this effect was significantly smaller in the conditions that induced fewer errors. Responses were also slowed when the head noun denoted a collective; there were additive effects of head noun collectivity and local noun number on singular RT and on the proportion of singular responses, replicating the pattern obtained by Haskell and MacDonald (2003). Errors made in conditions with an intervening number attractor were not reliably faster or slower than the corresponding correct responses; across the four experiments, there was no consistent trend in either direction, and in no condition was the difference between errors and correct responses fully significant. Experiment 4 found that a non-intervening attractor induced errors at a rate similar to an intervening attractor, but that the effect of a non-intervening attractor on correct RT was much smaller. In addition, errors made in the presence of a non-intervening attractor were significantly slower than correct responses.

The results confirm the predictions laid out in the Introduction regarding intervening number attraction. An intervening attractor affects both the difficulty of the agreement computation process, as reflected in RT, and its ultimate outcome, as reflected in response proportions, with these two variables being tightly coupled. There is no compelling evidence that agreement computation is especially difficult in those cases in which an error is actually made. An adequate explanation of number attraction errors should account for these findings, and as noted in the Introduction, one way to do this is in the manner of the Eberhard et al. (2005) model, which claims that a num-

ber attractor affects a continuously-valued representation of subject number on every trial, with errors arising from a probabilistic decision process that takes this representation as input. Fig. 4 suggests that in fact $S(r)$, the continuously-valued quantity that the Eberhard et al. (2005) model takes response proportions to reflect, may actually be a better linear predictor of RT than are response proportions themselves.

The results of Experiment 4 suggest that non-intervening attraction should receive a different kind of explanation (as originally proposed by Bock & Miller, 1991), as it appears that a non-intervening attractor has its most dramatic effect on RT when an error is actually made. The findings from Experiment 4 are also critical in demonstrating the ability of the experimental paradigm used here to detect slow errors; in light of these findings, the contrasting results regarding intervening attraction are especially compelling.

As noted above, one interesting feature of an account like that of Eberhard et al. is that number attraction errors (when these are due to an intervening attractor) are not interpreted in representational terms. Unlike the percolation account, which claims that errors arise in those cases in which the wrong number feature has been associated with the subject phrase as a whole, the Eberhard et al. account asserts no representational difference between errors and correct agreement decisions. The percolation account fails to predict the results obtained here primarily because, at least on the version advanced in the literature, it would not seem to predict any RT effects at all. But is important to ask whether, more generally, it is possible for an account that assumes a representational difference between errors and correct agreement decisions to deal with the present results. The answer is clearly yes, with the caveat that such an account must assume that the speaker's final representation of subject number is settled through competition between activated representations, and that this competitive process takes time. For example, the present results could be predicted by a percolation-style account which claims that number features within the subject phrase engage in resource-demanding competition for instantiation on the root node of the phrase. In the terms of the Eberhard et al. model, one might imagine a process of time-consuming competition between the various $S(m)$ values in the subject phrase for influence of $S(r)$. In this variant, trial-to-trial variability in verb number might emerge from this competitive process of number valuation rather than from a stochastic decision process taking subject number as input.

In fact, a representational account that would seem to involve the appropriate competitive machinery has recently been proposed by Badecker and Kuminiak (2007), who suggested that attraction errors (both number attraction and, in their own studies, gender attraction) result from an error arising within a content-addressable retrieval mechanism (cf. Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006). They proposed that computing subject–verb agreement requires retrieval of the subject from working memory, presumably at a relatively late stage of the production process. This retrieval operation is cue-based rather than, e.g., serial, with the search being guided by various cues to subjecthood. Badecker and Kuminiak write that:

> [t]he working memory retrieval account of agreement processing locates attraction errors in competition among elements in working memory that retrieval mechanisms have identified as agreement sources. Contests between head and local nouns arise because the mechanism for retrieving the agreement source (cue-based retrieval) makes the agreement process susceptible to mis-identifying a source. The susceptibility is due at least in part to the fact that some local nouns will resonate to certain retrieval cues as well as the intended retrieval target (the head of the subject phrase). (2007, p. 82)

The retrieval account has means to deal with both the syntactic depth effect and the mismatch asymmetry, and some attraction phenomena that do not appear to have a natural explanation on the percolation account are more easily explained on the retrieval view. For example, case disambiguation of the nouns in the subject phrase, in German and Dutch, seems to eliminate attraction errors (Hartsuiker et al., 2003). Badecker and Kuminiak also found that overt case marking eliminated gender attraction errors in Slovak. Overt case marking would not obviously affect the probability of a local noun's number feature percolating up the structure of the subject phrase, but it is plausible that overt nominative case would be a valid subject retrieval cue, and that overtly non-nominative case would make a local noun less likely to be mis-identified as the subject. However, the retrieval account would also seem to have certain drawbacks. For example, it would seem to predict an increase in attraction errors when the local noun is an especially likely subject on semantic grounds. As noted above, Bock and Miller (1991) and Barker et al. (2001) found that in the standard intervening configuration, an animate attractor was not especially likely to induce an error.

Though this is inexplicit in Badecker and Kuminiak's exposition, it is easy to see how competition between elements in working memory could take time, and that the intensity of the competition (i.e., how evenly-matched the retrieval cues are) would predict both response proportions and the time it takes the retrieval process to complete. It is also plausible that errors and correct responses should have similar latencies, as the duration of the retrieval process would be due to the intensity of the competition between nouns, regardless of outcome.

The present data, then, cannot definitively arbitrate between an account according to which there is no representational difference between errors and correct responses (Eberhard et al., 2005), and an account on which there is such a difference, as long as this latter account assumes that the final representation of subject number is arrived at by way of a time-consuming competition process. However, there are at least two arguments in favor of the former account. The first is parsimony. The choice is between an account on which the probability of a singular or plural verb directly reflects the relative strength of singular and plural number information in the subject phrase, and an account on which overt agreement behavior in a

given instance reflects a determinate "inner" representation of subject number, which in turn reflects, in a probabilistic manner, the relative strength of singular and plural information. There is no obvious theoretical motivation for this two-stage model, if a one-stage model achieves equal empirical coverage. The one-stage model does away with an undetectable (and perhaps philosophically suspect) representational level.

Secondly, representational accounts such as the retrieval account would seem to make one important empirical prediction that is not borne out by the data. If number attraction errors are due to the speaker or comprehender actually being in error about which noun within the subject phrase is the agreement controller, then this error should have ramifications downstream. There are two studies bearing on this question. Antón-Méndez, Nicol, and Garrett (2002) took advantage of the fact that in Spanish a predicate adjective must agree with the subject in both gender and number. Their experiment varied both the number and the grammatical gender of the head and local nouns in a visual preamble, which was presented together with an adjective. The participant was required to produce a predicative sentence, which required producing both a number-inflected copula and selecting the appropriate number and gender inflections for the adjective. If number attraction is due to the speaker treating the wrong noun as the head of the subject phrase, then number agreement errors and gender agreement errors ought to occur together. In fact, though the paradigm successfully induced both number and gender attraction errors, these two types of errors were statistically independent: The errors occurred together only about as often as would be expected based on the product of the probabilities of each error type. (Number attraction errors on the copula and on the adjective did have a strong tendency to occur together, but this result would be expected based on the assumption that there is only a single decision process that determines predicate number.)

More recently, Lau, Wagers, Stroud, and Phillips (2008) performed a comprehension experiment, using self-paced reading, in which number agreement in a cleft construction was either correct or in error, and in which a verb–object relation was either plausible or implausible (e.g., *The phone by the toilet(s) were/was what Patrick dialed/flushed as he...*) As expected, they found a number attraction effect on the copula, as the reading time effect of an agreement violation was reduced when the local noun was plural. They also found that an implausible verb–object relation led to increased reading times just after the critical verb. However, there was no hint of an interaction of these manipulations, suggesting that attraction did not actually lead readers to take the local noun to be the subject. If readers had taken *toilets* to be the subject, then *flushed* would have been plausible, not implausible, and presumably the reading time effect of the implausible thematic relation would have been reduced.

In sum, based on the results of Antón-Méndez et al. (2002) it appears that when speakers make number attraction errors they do not also take the subject to have the local noun's grammatical gender, and based on the results of Lau et al. (2008) it appears that when readers fall victim to number attraction they do not also treat the local noun as the semantic subject. These results would seem difficult to square with the view that number attraction is due to a retrieval process identifying the wrong noun as the head of the subject phrase. It should be acknowledged that a retrieval-based account could make amendments to deal with these findings, perhaps by stipulating that the retrieval in question is specific to number agreement itself, and is not visible to the processing mechanism more generally, but it is not clear that such amendments would be well motivated.

Turning to a different issue, the present research suggests that intervening and non-intervening attraction may require different kinds of explanations. Eberhard et al. (2005) recognize that their phrase-internal spreading activation architecture cannot account for attraction from outside the subject phrase, and suggest that the model may ultimately have to place weight on syntactic connections between the root node of the subject phrase and nodes that are outside of this phrase. But in fact, the results of Experiment 4 suggest that, at least in the case of non-intervening attraction from the head of a relative clause, such an amendment may not be necessary, as the RT patterns associated with this type of attraction suggest that a rather different phenomenon may be at work. (There are other forms of attraction from outside the subject phrase, such as object attraction in languages with subject–object–verb order (Hartsuiker, Antón-Méndez, & van Zee, 2001) which may still need to be accounted for.)

Finally, some notes on methodology are in order. The present paradigm is clearly less naturalistic than the standard laboratory production paradigm for studying number attraction (which is itself not all that naturalistic), and it clearly involves elements of comprehension as well as production. Thus, it might be argued that the results obtained here are not informative about normal production. In response to this objection, it is important to emphasize the remarkable ability of this paradigm to mimic the standard production results. This paradigm showed the number attraction effect, the mismatch asymmetry, the syntactic depth effect, the tendency of collective heads to trigger plural agreement, and non-intervening attraction. Thus, an obvious inference is that the paradigm used here is indeed tapping into the aspects of the production process that are relevant to subject–verb agreement computation. Indeed, the fact that all these effects could be obtained in the present paradigm is itself informative, for this suggests that these effects probably do not arise at the stage of message-level planning of an utterance (as in the present paradigm participants never produced the subject phrase itself), but rather at a later stage of checking the verb's number feature.

Acknowledgments

This research was supported by National Institutes of Health Grant HD-18708 to the University of Massachusetts. Parts of this research contributed to the author's doctoral dissertation. Thanks to Chuck Clifton, Lyn Frazier, Keith Rayner, Andrew Cohen, Lisa Scott, and Meg Grant for helpful discussion and comments.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jml.2008.11.002.

References

Antón-Méndez, Inés, Nicol, Janet L., & Garrett, Merrill F. (2002). The relation between gender and number agreement processing. *Syntax, 5*, 1–25.

Baayen, R. H., Davidson, D. H., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement, and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language, 56*, 65–85.

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

Bates, D. M. (2005). Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project, 5*(1), 27–30.

Bock, J. K., Butterfield, S., Cutler, A., Cutting, J. C., Eberhard, K. M., & Humphreys, K. R. (2006). Number agreement in British and American English: Disagreeing to agree collectively. *Language, 82*, 64–113.

Bock, J. K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language, 31*, 99–127.

Bock, J. K., & Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Language and Cognitive Processes, 8*, 57–99.

Bock, J. K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language, 51*, 251–278.

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

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

Chomsky, N. (1995). *The mimimalist program*. Cambridge, MA: MIT Press.

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

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

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making sense of syntax: Number agreement in sentence production. *Psychological Review, 112*, 531–559.

Franck, J., Cronel-Ohayon, S., Chillier, L., Frauenfelder, U., Hamann, C., Rizzi, L., et al (2004). Normal and pathological development of subject–verb agreement in speech production: A study on French children. *Journal of Neurolinguistics, 17*, 147–180.

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition, 101*, 173–216.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject–verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes, 17*, 371–404.

Freedman, S. E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition, 19*, 101–131.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language, 58*, 161–187.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1411–1423.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*, 97–114.

Hartsuiker, R. J., Antón-Méndez, I., & van Zee, M. (2001). Object attraction in subject–verb agreement construction. *Journal of Memory and Language, 45*, 546–573.

Hartsuiker, R. J., Schriefers, H. J., Bock, J. K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject–verb agreement. *Memory and Cognition, 31*, 1316–1326.

Haskell, T. R., & MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *Journal of Memory and Language, 48*, 760–778.

Humphreys, K. R., & Bock, J. K. (2005). Notional number agreement in English. *Psychonomic Bulletin & Review, 12*, 689–695.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Kayne, R. (1989). Romance clitics, verb movement and PRO. *Linguistic Inquiry, 22*, 647–686.

Kimball, J., & Aissen, J. (1971). I think, you think, he think. *Linguistic Inquiry, 2*, 241–246.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602.

Lau, E., Wagers, M., Stroud, C., & Phillips, C. (2008). *Agreement and the subject of confusion*. Paper presented at 21st Annual CUNY Conference on Human Sentence Processing, Chapel Hill, NC.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science, 10*, 447–454.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*, 50–68.

Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language, 54*, 466–490.

Meyer, A. S., & Bock, J. K. (1999). Presentations and processes in the production of pronouns: Some perspectives from Dutch. *Journal of Memory and Language, 41*, 281–301.

Nicol, J. (1995). Effects of clausal structure on subject–verb agreement errors. *Journal of Psycholinguistic Research, 24*, 507–516.

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

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*, 266–300.

Nosofsky, R. M., & Stanton, R. D. (2005). Speeded classification in a probabilistic category structure: Contrasting exemplar-retrieval, decision-boundary, and prototype models. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 608–629.

Nosofsky, R. M., & Stanton, R. D. (2006). Speeded old–new recognition of multidimensional perceptual stimuli: Modeling performance at the individual-participant and individual-item levels. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 314–334.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 11*, 427–456.

Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. Keiras & M. Just (Eds.), *New methods in reading comprehension research* (pp. 91–118). Hillsdale, NJ: Erlbaum.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159–182.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*, 873–922.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333–367.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*, 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General, 133*, 63–82.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh: Psychology Software Tools, Inc.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology, 49*, 1–46.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–661.

Traxler, M., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47*, 69–90.

Traxler, M., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53*, 204–224.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject–verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language, 34*, 186–215.

Vigliocco, G., & 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., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language, 45*, 368–390.

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes, 11*, 407–442.

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition, 68*, 13–29.

Vigliocco, G., & Zilli, T. (1999). Syntactic accuracy in sentence production: The case of gender disagreement in Italian language-impaired and unimpaired speakers. *Journal of Psycholinguistic Research, 28*, 623–648.

Waters, G. S., & Caplan, D. (1996). Processing resource capacity and the comprehension of garden path sentences. *Memory and Cognition, 24*, 342–355.