# Teasing apart encoding and retrieval interference in sentence comprehension: Evidence from agreement attraction

**Dan Parker (dparker@wm.edu)**
Linguistics Program, William & Mary, P.O. Box 8795
Williamsburg, VA 23187 USA

**Kelly Konrad (kpkonrad@email.wm.edu)**
Linguistics Program, William & Mary, P.O. Box 8795
Williamsburg, VA 23187 USA

**Abstract**

This study investigates interference effects in sentence processing. A parade case involves agreement attraction, where the processing of a number mismatch between a verb and its subject is eased by a number-matching lure (**The key*$_{\text{target}}$ *to the cabinets*$_{\text{lure}}$ *were rusty*), relative to sentences where neither noun matches the verb (**The key to the cabinet were rusty*). Existing accounts claim that this effect reflects error-prone retrieval or misrepresentation of the target. Recently, a third account has been proposed which claims that the contrast between the two configurations reflects increased difficulty in the second sentence due to feature overwriting in the encoding (both nouns are singular). We provide results from two self-paced reading experiments that isolate the effects of feature overwriting and attraction by manipulating the presence of an agreement cue. Results showed a larger difference within the configurations with a cue, which suggest that attraction cannot be reduced to feature overwriting.

**Keywords:** sentence processing, interference, agreement attraction, memory retrieval, feature overwriting, reading times

## Introduction

Interference effects provide valuable clues about how we mentally encode and access linguistic information in working memory during language processing (Jäger, Engelmann, & Vasishth, 2017). One type of interference effect that has received much attention in the sentence comprehension literature involves so-called "agreement attraction" (Clifton, Frazier, & Deevy, 1999; Pearlmutter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2009), where the processing of a number mismatch between a verb and its subject is eased by a number-matching lure noun (the "attractor"), e.g., (1a), relative to equally ungrammatical sentences that lack a number-matching noun, e.g., (1b).

(1) a. *The key$_{\text{(target)}}$ to **the cabinets**$_{\text{(lure)}}$ unsurprisingly **were** rusty.
b. *The key to the cabinet unsurprisingly **were** rusty.

There are two leading accounts of agreement attraction. One account pins the problem on error-prone memory retrieval mechanisms (Wagers et al., 2009). On this account, the plural verb *were* in (1) triggers a retrieval process to recover an item in memory that matches the cues [+subject] and [+plural]. In (1a), this process may erroneously retrieve the plural lure, e.g., *the cabinets*, based on the partial match to the [+plural] cue, leading to the false impression that agreement is licensed (see also Dillon, Mishler, Sloggett, & Phillips, 2013; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Tanner, Nicol, & Brehm, 2014; Tucker, Idrissi, & Almeida, 2015). A competing account suggests that attraction reflects misrepresentation of the target subject, rather than misretrieval (see Hammerly, Staub, & Dillon, 2019, for a review). One version of this account claims that the plural feature of the attractor "percolates" up to the target subject, spuriously licensing agreement (Bock & Eberhard, 1993; Eberhard, 1997; Franck, Vigliocco, & Nicol, 2002; Vigliocco, Butterworth, & Semenza, 1995). Another version claims that spreading activation of the plural number on the attractor triggers agreement (Eberhard, Cutting, & Bock, 2005).

Recently, a third account has been proposed. Vasishth and colleagues (Vasishth, Jäger, & Nicenboim, 2017) argued that the contrast between the sentences in (1) does not reflect misretrieval or misrepresentation of the target, but rather increased processing difficulty in (1b), relative to (1a), due to feature overwriting at the stage of encoding. Vasishth et al. point out that whereas the nouns in the attractor-match condition (1a) have different number markings (target = singular, attractor = plural), the nouns in the attractor-mismatch condition (1b) are both singular. In this scenario, a process known as "feature overwriting" (Nairne, 1990), can occur, in which the number feature shared between the items becomes degraded, making retrieval of the target (i.e., *the key*) more difficult. This effect constitutes a form of interference whereby overlap in features between the target and a lure deteriorates the quality of their representations in memory, which impedes access to the target, predicting processing difficulty in the form of a slow down at the point of retrieval, e.g., at the verb.

The feature overwriting account is attractive because it does not require stipulation of any new mechanisms and Vasishth et al. (2017) offer an explicit computational model of their account that provides a good fit to existing data. However, their account misses a key point about agreement attraction: comprehenders find the attractor-match condition (1a) to be on a par with grammatical agreement, e.g., *The key to the cabinet(s) is rusty*, giving rise to an "illusion of acceptability" (Phillips, Wagers, & Lau, 2011). Crucially, the

feature overwriting account does not explain why comprehenders are fooled into accepting (1a).

It is difficult to distinguish the competing accounts of agreement attraction because the data (e.g., reading times, acceptability judgments) underdetermines the underlying generative processes, i.e., there are multiple cognitive processes that could give rise to the observed behavior. Here, we set out to test the predictions of the feature overwriting account. Specifically, if the contrast between (1a) and (1b) reflects increased processing difficulty due to feature overwriting in the attractor-mismatch condition (1b), then the same contrast should arise even when the verb does not deploy a number cue for retrieval, as in the case with past tense verbs, e.g., *The key to the cabinets apparently* ***had been misplaced***.

This prediction was first extrapolated in Villata et al. (2018). Villata and colleagues tested retrieval for agreement processing in configurations with and without a number cue like in Table 1 and found that overlap in number features between the target (e.g., *the waiter*) and lure (e.g., *the dancer(s)*) had a marginal effect on agreement processing, but only when retrieval required number agreement (e.g., *criticizes*). Villata and colleagues presented their findings as evidence for interference at retrieval.

Table 1: Sample items from Villata et al. (2018) Expt. 2.

| | |
|---|---|
| **+cue, +match** | The dancer-SG that the waiter-SG strongly criticizes-SG … |
| **+cue, -match** | The dancers-PL that the waiter-SG strongly criticizes-SG … |
| **-cue, +match** | The dancer-SG that the waiter-SG strongly criticized-Ø … |
| **-cue, -match** | The dancers-PL that the waiter-SG strongly criticized-Ø … |

There are two reasons to revisit the claims in Villata et al. (2018). First, they did not test the critical agreement attraction configuration in (1a), focusing instead on grammatical sentences where the target matched the verb in number. Importantly, a growing number of studies suggest that subject-verb agreement is computed differently in grammatical and ungrammatical configurations (Lago, Alcocer, & Phillips, 2011; Wagers et al., 2009). For instance, in grammatical configurations, agreement can be computed via predictive processing, e.g., the target subject predicts the number of the verb. However, when the verb form violates this prediction in ungrammatical configurations like those in (1), comprehenders engage memory retrieval as a repair/reanalysis procedure to recover a number matching item to license agreement. This difference might explain why Villata et al. (2018) did not find a significant effect of retrieval in grammatical contexts. But more research is needed on the configurations in (1), which are argued to engage retrieval. Second, singular verbs like those in the +cue conditions of their study generally do not induce interference effects. Instead, research shows that only plural verbs trigger interference, resulting in a "plural markedness effect" (see Wagers et al., 2009, for discussion). That is, the conditions that Villata et al. may not have been an appropriate test for interference effects. It thus remains unclear whether interference in configurations like (1a) can be reduced to feature overwriting in the encoding.

## Experiment 1

Experiment 1 tests the predictions of the feature overwriting account by extending the design developed by Villata et al. (2018) to configurations that trigger agreement attraction. Specifically, we used a 2 × 2 design that manipulated (i) the number overlap between the target subject and a PP attractor (overlap vs. no overlap), and (ii) the presence of an agreement cue on the verb (+cue vs. -cue), as shown in Table 2. This design isolates the effect of feature overwriting with the -cue conditions, allowing us to compare the profile of feature overwriting to the attraction effect in the +cue conditions. If agreement attraction really reflects feature overwriting, we should see comparable differences within the +cue and -cue conditions. If attraction has a different underlying process (e.g., misretrieval, feature misrepresentation), then we see should a greater difference within the +cue conditions, above and beyond any effect of feature overwriting revealed in the -cue conditions.

Importantly, this design overcomes the two main issues concerning the original design tested by Villata et al. (2018). First, retrieval is required in both the +cue and -cue conditions. It is assumed that in the +cue conditions, retrieval is engaged at the verb in response to the number prediction error generated by the subject (Wagers et al., 2009). In the -cue conditions, retrieval is required to relate the subject and verb thematically. Second, the design employs plural verbs in the +cue conditions, which reliably induce interference effects (Wagers et al., 2009).

Table 2: Sample items from Experiment 1.

| | |
|---|---|
| **-overlap, +cue** | The key to the cabinets apparently have been misplaced by the guard. |
| **+overlap, +cue** | The key to the cabinet apparently have been misplaced by the guard. |
| **-overlap, -cue** | The key to the cabinets apparently had been misplaced by the guard. |
| **+overlap, -cue** | The key to the cabinet apparently had been misplaced by the guard. |

## Participants

Participants were 120 native speakers of English recruited from [author's institution]. Participants received credit in an introductory psychology or linguistics course. The experiment session lasted approximately 25 min.

## Materials

Experimental materials were harvested from Wagers et al. (2009) and modified to create 24 sets of 4 conditions, as shown in Table 1. Across all item sets, the target subject (e.g., *the key*) was modified by a prepositional phrase that contained the attractor (e.g., *cabinets*). Number overlap between the target and attractor was manipulated by varying the number of the attractor (singular/plural) to either match or mismatch the singular target NP. The critical auxiliary verb was always a form of *has*: the +cue conditions used *have*, which required number agreement, and the -cue conditions used past tense *had*, which does not require number agreement.[1]

The 24 target items were distributed across 4 lists in a Latin square design and combined with 48 grammatical filler sentences of similar length and complexity, such that each participant read a total of 72 sentences. All sentences were followed by a 'yes/no' comprehension question that addressed various parts of the sentence to prevent participants from developing superficial reading strategies that would allow them to answer the question without reading the entire sentence.

## Procedure

The experiment was conducted using the online experiment platform Ibex (Drummond, n.d.), which allows self-paced reading experiments to be deployed in a standard web browser. Sentences were initially masked by dashes, with white spaces and punctuation intact. Participants pushed the space bar to reveal each word. Presentation was non-cumulative, such that the previous word was replaced with dashes when the next word appeared. On-screen feedback was provided for incorrect answers to the comprehension questions. The order of presentation was randomized for each participant.

## Analysis

Data from all participants were included in the analysis. Statistical analyses were carried out over the untrimmed, log-transformed reading time data with linear mixed-effects models using the *lme4* package (Bates, Maechler, & Bolker, 2011) in the *R* software environment (R Development Core Team, 2020). Models were defined using orthogonal contrast coding to examine the effects of number overlap, number cue, and their interaction (overlap × cue) for three regions of interest, including the critical auxiliary verb (critical region) and the following two words (spillover regions 1 and 2). All models were fit with a full variance-covariance matrix, i.e., a maximal random effects structure, with random intercepts and slopes for all fixed effect predictors by participants and items (Barr, Levy, Scheepers, & Tily, 2014). If there was a convergence failure or if the model converged but the

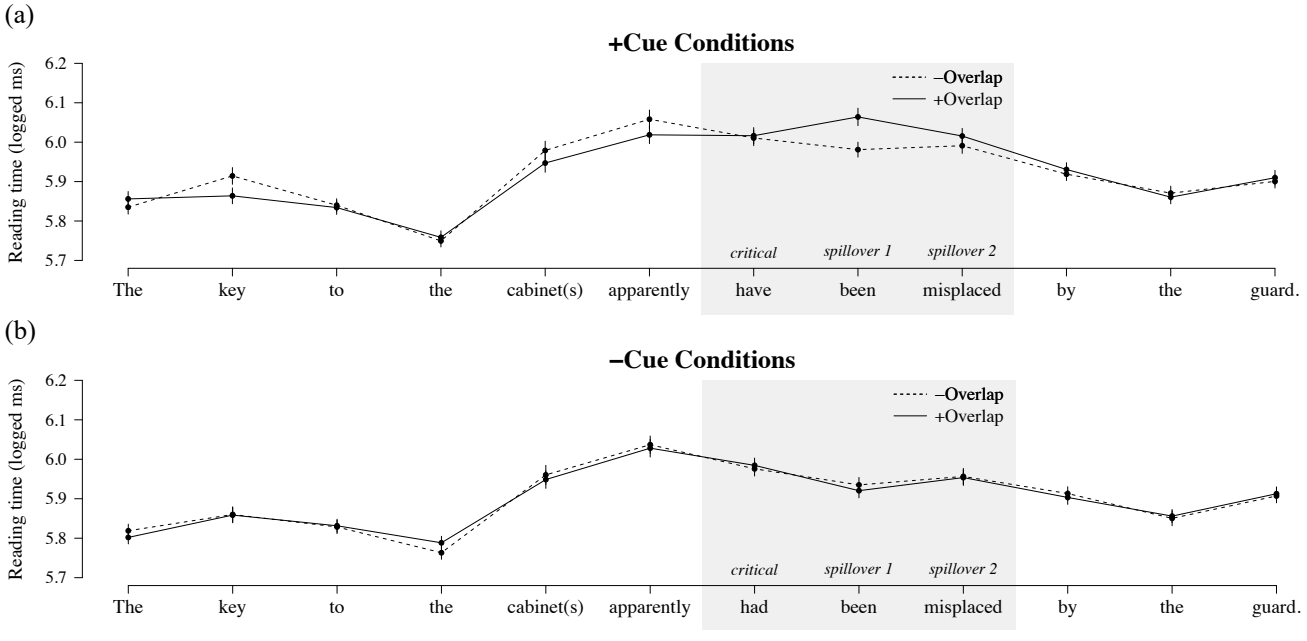


Figure 1. Word-by-word reading times for the +cue (a) and -cue (b) conditions in Experiment 1. Error bars indicate standard error of the man.

[1] All items, code, and data for this study are available on the Open Science Framework (https://osf.io/unk94/).

correlation estimates were high, the random effects structure was simplified. A fixed effect was considered significant if its absolute *t*-value was greater than 2, which indicates that its 95% confidence interval did not include 0 (Gelman & Hill, 2007).

### Results

Figure 1 shows the average word-by-word reading times for the four experimental conditions in Table 1. No effects were observed at the critical region (number overlap: $\hat{\beta} = 0.00$, SE = 0.02, $t$ = 0.30; cue: $\hat{\beta}$ = -0.03, SE = 0.02, $t$ = -1.71; interaction: $\hat{\beta} = 0.00$, SE = 0.02, $t = 0.16$). Spillover region 1 showed a main effect of number overlap ($\hat{\beta} = 0.08$, SE = 0.01, $t$ = 4.29), cue ($\hat{\beta}$ = -0.04, SE = 0.02, $t$ = -2.20) and an interaction of number overlap with cue ($\hat{\beta} = -0.09$, SE = 0.02, $t = -3.53$), driven by the difference in the +cue conditions ($\hat{\beta}$ = 0.08, SE = 0.02, $t$ = 3.99). No effects were observed in spillover region 2 (number overlap: $\hat{\beta} = 0.02$, SE = 0.02, $t$ = 1.19; cue: $\hat{\beta}$ = -0.03, SE = 0.02, $t$ = -1.61; interaction: $\hat{\beta}$ = -0.02, SE = 0.02, $t$ = -0.94).

### Discussion

Experiment 1 isolated the effect of feature overwriting and compared it to the effect of agreement attraction to better understand the source of agreement attraction effects in sentence comprehension. Specifically, Experiment 1 manipulated the number overlap between the target subject and a PP attractor (overlap vs. no overlap) and the presence of an agreement cue (+cue vs. -cue) on the verb. Results showed a larger difference (i.e., attraction effect) within the +cue conditions, above and beyond any effect of feature overwriting revealed in the -cue conditions. These results are incompatible a feature overwriting account of agreement attraction, which predicts that comparable effects should be observed within the +cue and -cue conditions.

A post-hoc analysis of the second NP region (i.e., the lure) suggested by an anonymous reviewer shows no effects (all *t*s < 1.45). Crucially, the features of the target NP are overwritten due to similarity with the lure, we might expect a reading time penalty at the lure region. However, the lack of any evidence for such an effect might be taken as additional evidence against the feature overwriting account.

One concern with Experiment 1 is that in the -cue conditions, the past tense auxiliary verb (*had*) might not have triggered retrieval. For instance, if the auxiliary does not require agreement, the parser might delay retrieval for subject-verb binding until the main verb (e.g., *had been **misplaced***) is encountered. Although there is not a statistically significant difference between the -cue conditions at or following the main verb (post-hoc analysis: *t*s < 2), it is important to keep the retrieval trigger in the same linear position across conditions to avoid a confound due to distance between the retrieval trigger and target item. This issue is addressed in Experiment 2.

## Experiment 2

The goal of Experiment 2 was to provide a conceptual replication of Experiment 1, holding constant the position of the retrieval trigger across conditions. To achieve this, Experiment 2 used predicates with full lexical verbs in place of auxiliary verbs as the retrieval trigger.

### Participants

Participants were 120 native speakers of English who were recruited using Amazon's Mechanical Turk web service (https://aws.amazon.com/mturk). All participants provided informed consent and were screened for native speaker abilities. The screening probed knowledge of the constraints on English tense, modality, morphology, ellipsis, and syntactic islands. Participants were compensated $3.00. The experiment lasted approximately 30 min.

### Materials

Experimental materials consisted of the same 24 sets of 4 conditions as in Experiment 1, with the same filler sentences. To keep the retrieval trigger constant across conditions, Experiment 2 used full lexical verbs, rather than auxiliary verbs, as the retrieval trigger, as shown in Table 2.

Table 2: Sample items from Experiment

| | |
|---|---|
| **-overlap, +cue** | The key to the cabinets apparently rust due to years of disuse. |
| **+overlap, +cue** | The key to the cabinet apparently rust due to years of disuse. |
| **-overlap, -cue** | The key to the cabinets apparently rusted due to years of disuse. |
| **+overlap, -cue** | The key to the cabinet apparently rusted due to years of disuse. |

### Procedure and analysis

Experiment 2 used self-paced reading, following the same procedure used in Experiment 1. Since the experiment was conducted remotely using Mechanical Turk, we employed an instructional manipulation check (Oppenheimer, Meyvis, & Davidenko, 2009) as an additional step to ensure that participants completed the task as directed. Instructional manipulation checks ensure that participants complete the task as directed by asking them to ignore the standard response format and provide a confirmation that they have read the instructions.

Data analysis followed the same steps as in Experiment 1. Five participants were removed from the analysis for failing the instructional manipulation check, leaving data from 115 participants for the final analysis.

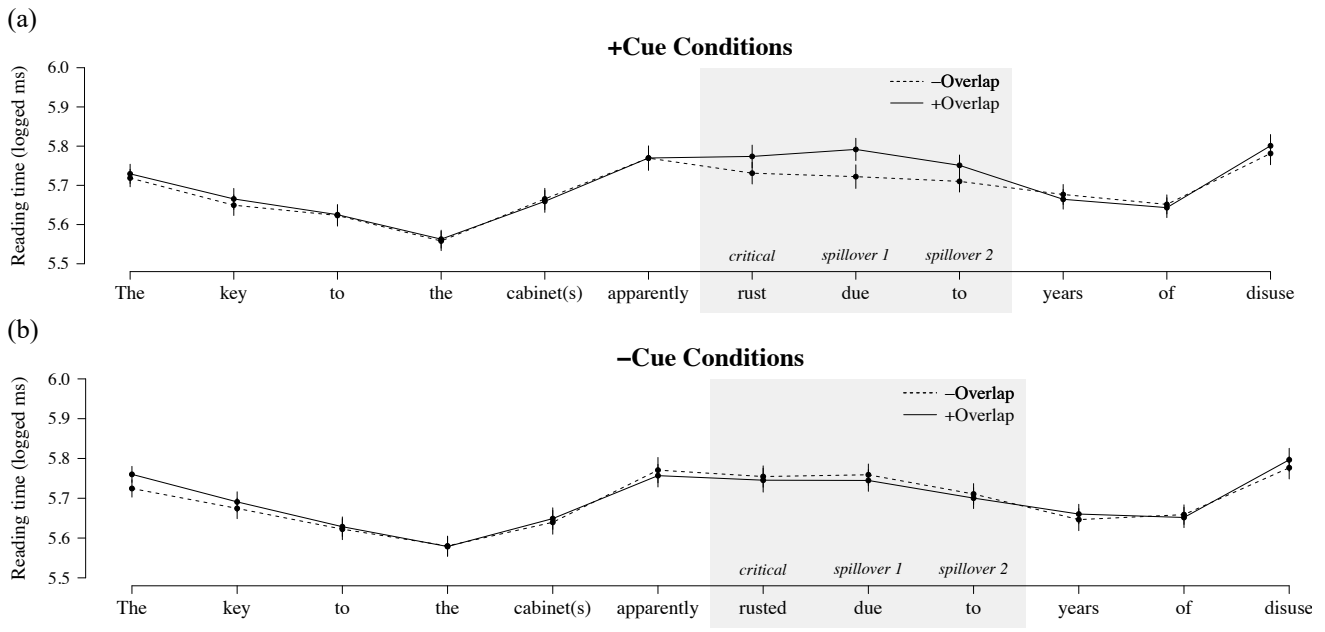


Figure 2. Word-by-word reading times for the +cue (a) and -cue (b) conditions in Experiment 2. Error bars indicate standard error of the man.

## Results

Figure 2 shows the average word-by-word reading times for the four experimental conditions in Table 2. No effects were observed at the critical region (number overlap: $\hat{\beta} = 0.02$, SE = 0.02, $t = 1.06$; cue: $\hat{\beta} = 0.04$, SE = 0.02, $t = 1.96$; interaction: $\hat{\beta} = -0.05$, SE = 0.03, $t = -1.67$). Spillover region 1 showed a main effect cue ($\hat{\beta} = -0.06$, SE = 0.02, $t = 2.91$) and an interaction of number overlap with cue ($\hat{\beta} = -0.08$, SE = 0.03, $t = -2.53$), driven by the difference in the +cue conditions ($\hat{\beta} = 0.06$, SE = 0.02, $t = 2.70$). No effects were observed in spillover region 2 (number overlap: $\hat{\beta} = 0.00$, SE = 0.02, $t = 0.03$; cue: $\hat{\beta} = 0.04$, SE = 0.02, $t = 1.93$; interaction: $\hat{\beta} = -0.05$, SE = 0.02, $t = -1.72$).

## Discussion

Experiment 2 provided a conceptual replication of Experiment 1 using full lexical verbs as the retrieval trigger in place of auxiliary verbs to control for effects of distance. As in Experiment 1, results showed a difference within the conditions with an agreement retrieval cue and beyond any effect of feature overwriting revealed in the conditions without an agreement cue. Taken together, these results suggest that agreement attraction cannot be reduced to feature overwriting at the stage of encoding.

# General Discussion

The goal of the current study was to better understand the underlying generative process that gives rise to agreement attraction effects in sentence comprehension. Such effects are typically characterized as eased processing of a subject-verb number mismatch in the presence of a number-matching attractor, relative to sentences that lack a number-matching NP. Previously, such effects have been attributed to error-prone memory retrieval mechanisms or misrepresentation of the target subject. Recently, a third account has been introduced which claims that the contrast between the conditions with and without a number-matching attractor actually reflects feature overwriting in the condition that lacks a number-matching attractor. On this account, feature similarity between the candidate agreement controllers degrades the quality of the target representation in memory, making it more difficult to recover the target later at retrieval, giving rise to the timing difference in previous studies.

To test this proposal, we isolated the effect of feature overwriting by controlling for the use of a number retrieval cue on the verb and compared the effect to that observed in agreement attraction configurations. Results from Experiments 1 and 2 both showed a larger difference (i.e., attraction effect) within the +cue conditions, above and beyond any effect of feature overwriting revealed in the -cue conditions, as shown in Figure 3. These results suggest that the observed reading time differences observed in previous tests of agreement attraction cannot be reduced to feature overwriting at the stage of the encoding.

A concern with the current study raised by an anonymous reviewer is that grammaticality and the presence of a number cue are confounded in Experiments 1 and 2: the +cue conditions are ungrammatical, whereas the -cue conditions are grammatical. Although we assumed that retrieval occurs in both the +cue and -cue conditions, the trigger for retrieval differs in these cases, e.g., prediction error vs. subject-verb thematic linking. This difference might impact agreement processing, but it is unclear how or in what direction. Future

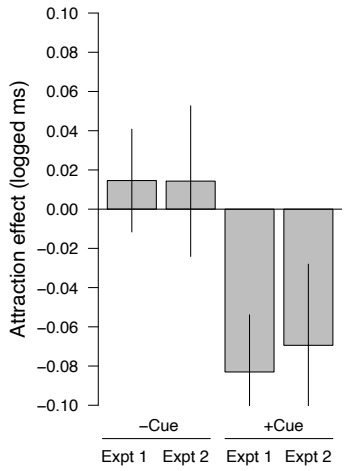


Figure 3. Comparison of effects for the +cue and -cue conditions across Experiments 1 and 2. Error bars indicate standard error of the mean

work should employ a design in which grammaticality is kept constant across the ±cue conditions.

Lastly, it is important to emphasize that the current results do not arbitrate between the retrieval and misrepresentation accounts. But they do underscore the importance of understanding the primary effect under investigation: agreement attraction leads to an *illusion of grammaticality* (Phillips et al., 2011), whereby ungrammatical conditions are processed on a par with the grammatical conditions. Crucially, the feature overwriting account does not explain this aspect of the phenomenon, and the results of the current study provide some empirical evidence that narrows down the space of possibilities by ruling out the feature overwriting account.

## Acknowledgments

We would like to thank the three anonymous reviewers and the members of the Computational & Experimental Linguistics Lab at William & Mary for their helpful feedback on this study. This work was supported in part by NSF grant BCS-18843309 awarded to Dan Parker.

## References

Barr, D., Levy, R., Scheepers, C., & Tily, H. J. (2014). Keep it maximal. *Journal of Memory and Language*, *68*(3), 1–43.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes.

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in english number agreement. *Language and Cognitive Processes*, *8*(1), 57–99.

Clifton, C. J., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista Di Linguistica*, *11*, 11–39.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Lanuage*, *69*, 85–103.

Drummond, A. (n.d.). Ibex Farm.

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

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

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.

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. In *Analytical methods for social research*.

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, *110*(January), 70–104.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 305–315.

Lago, S., Alcocer, P., & Phillips, C. (2011). *Agreement attraction in Spanish: Immediate vs. delayed sensitivity*.

Lago, S., Shalom, D., Sigman, M., Lau, E., & Phillips, C. (2015). Agreement processes in Spanish comprehension. *Journal of Memory and Language*.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872.

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

Phillips, C., Wagers, M., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the Interfaces* (Vol. 37, pp. 147–180).

R Development Core Team. (2020). *R: A language and environment for statistical computing*.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215.

Tucker, M. A., Idrissi, A., & Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology*, *6*.

Vasishth, S., Jäger, L. A., & Nicenboim, B. (2017). Feature overwriting as a finite mixture process: Evidence from

comprehension data. *MathPsych/ICCM*, (2015).
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*.
Villata, S., Tabor, W., & Franck, J. (2018). Encoding and Retrieval Interference in Sentence Comprehension: Evidence from Agreement. *Frontiers in Psychology*, *9*, 1–16.
Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.