

# UC Merced

## Proceedings of the Annual Meeting of the Cognitive Science Society

### Title

Feature encoding modulates cue-based retrieval: Modeling interference effects in both grammatical and ungrammatical sentences

### Permalink

<https://escholarship.org/uc/item/75k9f50t>

### Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

### Authors

Yadav, Himanshu

Smith, Garrett

Vasishth, Shravan

### Publication Date

2021

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Feature encoding modulates cue-based retrieval: Modeling interference effects in both grammatical and ungrammatical sentences

Himanshu Yadav (hyadav@uni-potsdam.de)

Department of Linguistics, University of Potsdam  
14476 Potsdam, Germany

Garrett Smith (gasmith@uni-potsdam.de)

Department of Linguistics, University of Potsdam  
14476 Potsdam, Germany

Shravan Vasishth (vasishth@uni-potsdam.de)

Department of Linguistics, University of Potsdam  
14476 Potsdam, Germany

## Abstract

Studies on similarity-based interference in subject-verb number agreement dependencies have found a consistent facilitatory effect in ungrammatical sentences but no conclusive effect in grammatical sentences. Existing models propose that interference is caused either by a faulty representation of the input (encoding-based models) or by difficulty in retrieving the subject based on cues at the verb (retrieval-based models). Neither class of model captures the observed patterns in human reading time data. We propose a new model that integrates a feature encoding mechanism into an existing cue-based retrieval model. Our model outperforms the cue-based retrieval model in explaining interference effect data from both grammatical and ungrammatical sentences. These modeling results yield a new insight into sentence processing, encoding modulates retrieval. Nouns stored in memory undergo feature distortion, which in turn affects how retrieval unfolds during dependency completion.

**Keywords:** Similarity-based interference; encoding interference; cue-based retrieval

## Introduction

A well-studied phenomenon in sentence processing is similarity-based interference in subject-verb number agreement dependencies (Dillon, Mishler, Sloggett, & Phillips, 2013; Wagers, Lau, & Phillips, 2009; Lago, Shalom, Sigman, Lau, & Phillips, 2015). In this paradigm, a noun phrase matching in number with the verb—called a distractor—is presented along with the subject noun phrase. For example, in the following conditions (a) and (c), *the cabinets* matches the plural feature of the verb in contrast to conditions (b) and (d), in which the number feature of the subject differs from that of the verb.

### (a) Grammatical, interference condition

The key to the cabinet unsurprisingly was rusty.

### (b) Grammatical, no-interference condition

The key to the cabinets unsurprisingly was rusty.

### (c) Ungrammatical, interference condition

\* The key to the cabinets unsurprisingly were rusty.

### (d) Ungrammatical, no-interference condition

\* The key to the cabinet unsurprisingly were rusty.

A typical finding is facilitation in the ungrammatical conditions: reading times at the verb ‘were’ in condition (c) are faster than in condition (d) (Jäger, Engelmann, & Vasishth, 2017; Wagers et al., 2009; Lago et al., 2015; Dillon et al., 2013; Jäger, Mertzen, Van Dyke, & Vasishth, 2020). By contrast, there is no conclusive effect at the verb in the grammatical conditions: the effect ranges from a slowdown in condition (a) to speedup in condition (a) relative to condition (b). Figure 1 shows the pattern of interference effects in the grammatical and ungrammatical conditions from 11 published datasets.

Several attempts have been made to model interference effect in the subject-verb number agreement dependencies. While these models correctly predict facilitatory interference in the ungrammatical conditions, none of them is able to capture the observed interference effect patterns in the grammatical conditions.

Most previous models can be placed into one of two categories, encoding-based accounts and cue-based retrieval accounts. One encoding-based account (Bock & Eberhard, 1993; Eberhard, 1997; Vigliocco, Butterworth, & Semenza, 1995) assumes that the number feature of the distractor noun phrase percolates to the subject noun phrase causing a *misrepresentation* of the subject. For example, in condition (c), the plural feature of *the cabinets* percolates up to the subject *the key* in a proportion of trials. This enables the subject of the sentence to license a plural verb, which leads to faster dependency completion in condition (c) compared to condition (d), where no feature percolation happens. Thus, the encoding-based account correctly predicts the facilitatory effect in the ungrammatical conditions. In the grammatical conditions, the percolation of the plural feature to the subject in condition (b) will disrupt licensing of the verb in some proportion of trials. As a result, slower reading times are predicted in condition (b) compared to (a). This prediction is not supported by the interference effect data in grammatical conditions (see Fig. 1).

The cue-based retrieval account (Lewis & Vasishth, 2005;

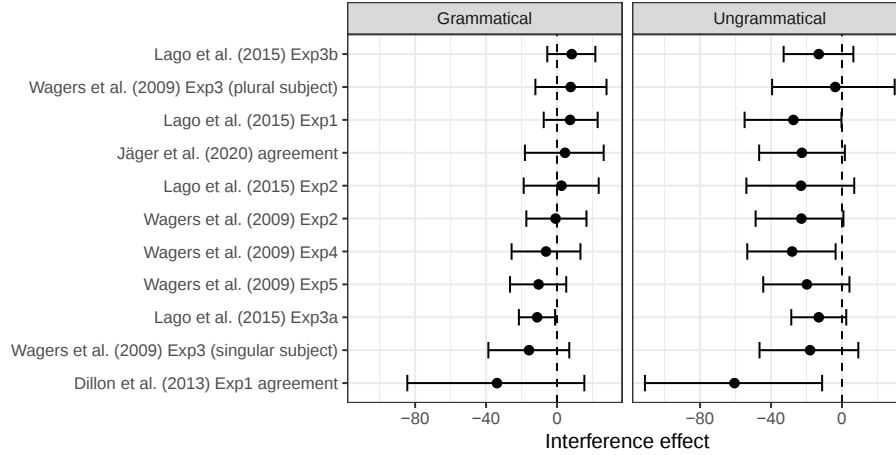


Figure 1: The pattern of interference effect in subject-verb agreement dependencies from four studies: Wagers et al. (2009), Dillon et al. (2013), Lago et al. (2015), and Jäger et al. (2020). Here, interference effect means the difference in reading times at the verb between the interference and no-interference conditions.

Engelmann, Jäger, & Vasishth, 2020) assumes that encountering a verb triggers a search in working memory based on the dependency completion requirements of the verb called retrieval cues. For example, in conditions (c) and (d), the verb requires a plural noun phrase in the subject position to complete a subject-verb dependency. Each noun phrase that matches a retrieval cue receives a certain amount of activation, and the chunk with the highest activation gets retrieved for dependency completion. In condition (c), each noun phrase, *the key* and *the cabinets*, partially matches the retrieval cues at the verb. Consequently, a race for retrieval is initiated such that either noun phrase can be retrieved for dependency completion. This causes statistical facilitation (Raab, 1962) in retrieval times in condition (c) compared to condition (d), where only one noun phrase matches the retrieval cues. Hence, cue-based retrieval correctly predicts the facilitatory effect in the ungrammatical conditions. In grammatical conditions, the model predicts a slowdown in condition (a) because the singular distractor noun reduces the total amount of activation to be received by the subject noun phrase compared to that in condition (b). Again, this prediction is not supported by the interference effect data in the grammatical conditions shown in Fig. 1.

To summarize, both encoding-based and retrieval-based models correctly predict the facilitatory effect in the ungrammatical conditions, but they fail to explain the observed range of effects in the grammatical conditions.

In this paper, we propose a new model of similarity-based interference to capture the interference effect patterns in both grammatical and ungrammatical conditions using a single set of mechanisms. We compare the performance of our model against the cue-based retrieval model of Lewis and Vasishth (2005) on interference effect data from 11 datasets shown in figure 1.

We first present the assumptions and basic outline of our

model. Next, we evaluate the predictions of the model and quantify evidence compared to the cue-based retrieval model. We then discuss the broader implications of the work and conclude.

### New model of interference effects

Our proposed model unifies feature encoding mechanisms and the cue-based retrieval process. We modify the existing cue-based retrieval model of Lewis and Vasishth (2005) such that the representation of the subject noun changes due to influence from the distractor noun before the retrieval is triggered at the verb. We call the new model the **encoding-plus-retrieval model**.

#### Assumptions

The encoding-plus-retrieval model assumes that

1. Dependency completion between the subject and the verb is driven by a cue-based retrieval process
2. Before retrieval occurs at the verb, probabilistic feature percolation occurs from the distractor noun to the subject noun

Before the retrieval is triggered at the verb, the representation of the subject noun changes in a proportion of trials due to percolation of the plural feature from the distractor to the subject noun. For example, in condition (b), the plural feature of *the cabinets* percolates up to the subject *the key* with a probability  $\theta$  (see figure 2). Consequently, out of total  $N$  trials, in  $\theta * N$  trials, the subject is misrepresented as a plural subject noun. When the retrieval is triggered in these trials, the subject receives activation only through the subject cue. As a result, the retrieval becomes slower compared to condition (a), where the subject NP receives activation through both subject and number cues (see figure 2).

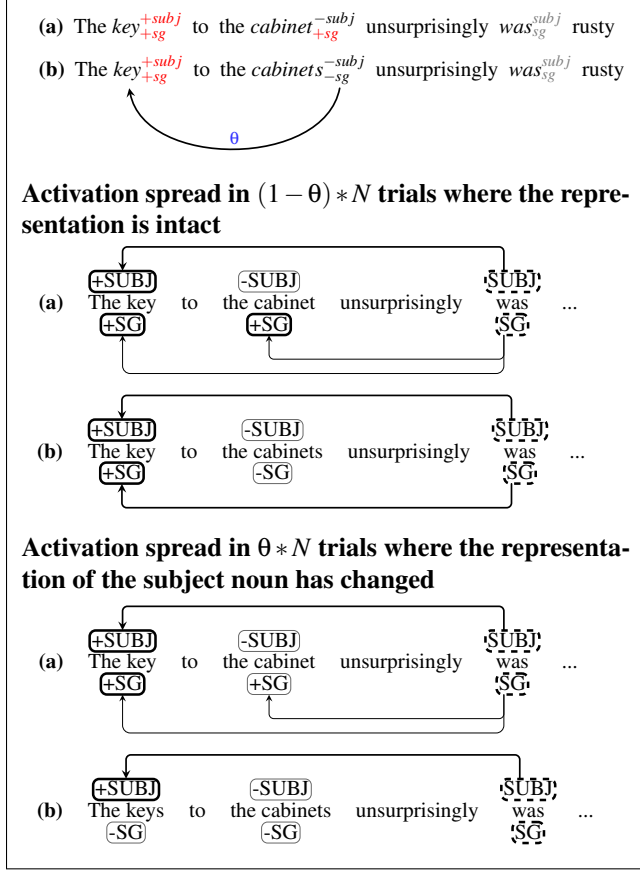


Figure 2: Feature percolation modulates the amount of activation received by the subject noun during the retrieval process. In  $\theta * N$  trials, the subject noun receives less activation in condition (b) causing a slowdown in retrieval times compared to condition (a).

Figure 3 shows that the model predicts that the percolation rate  $\theta$  modulates the sign of the interference effect in the grammatical conditions. As  $\theta$  increases, the effect changes from a slowdown in condition (a) to comparable reading times in (a) and (b) to a speedup in condition (a).

When the percolation rate is zero, the representation remains intact across all trials, and the subject noun receives less activation in condition (a) due to the presence of singular distractor noun compared to (b). As a result, the retrieval times in condition (a) are on average slower than in condition (b). As the percolation rate increases, the subject noun receives less and less activation in condition (b) compared to (a) (averaging across trials). Hence, the retrieval times in condition (a) become on average faster than in condition (b).

In the ungrammatical conditions, the effect changes quantitatively with an increase in  $\theta$ , but the effect is always facilitatory, i.e., faster reading times in the interference condition, (c) compared to no-interference condition, (d).

### Calculating retrieval times

Here, we derive the updated retrieval time equation to account for representation change due to feature percolation. Suppose

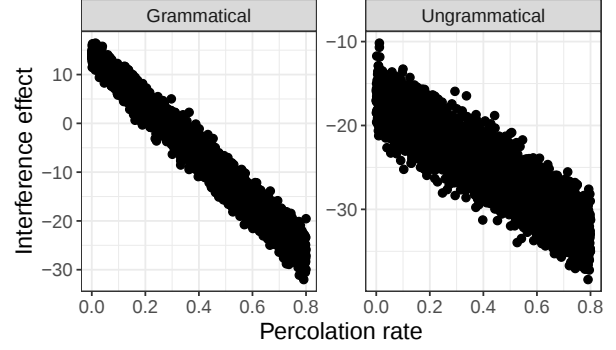


Figure 3: The interference effect as predicted by percolation rate,  $\theta$ . Negative values mean that the retrieval times in the interference condition are faster than the no-interference condition; positive values mean that the retrieval times in interference condition are slower than the no-interference condition.

$i$  indexes trial ID;  $N$  represents the total number of trials. The average retrieval time at the verb,  $RT$ , is given by,

$$RT = \frac{1}{N} \sum_{i=1}^N RT_i \quad (1)$$

where  $RT_i$  is the retrieval time in the  $i^{th}$  trial.

The retrieval time in the  $i^{th}$  trial is an exponential function of the activation of the retrieved chunk,

$$RT_i = Fe^{-A_{i, retrieved}} \quad (2)$$

where  $F$  is a scaling parameter called the latency factor.

The cue-based retrieval theory assumes that the chunk with the highest activation gets retrieved in each trial. The activation of the chunk retrieved in  $i^{th}$  trial would be the maximum of the activation of the subject and the distractor noun:

$$A_{i, retrieved} = \max(A_{i, subject}, A_{i, distractor}) \quad (3)$$

The activation of the subject and the distractor in a trial is determined by the amount of activation received by them due to cue-feature match. The noun phrase that matches more number of cues receives higher activation. Thus, activation of the subject in  $i^{th}$  trial,  $A_{i, subject}$  is a function of its representation,

$$A_{i, subject} = \text{Activation}(R_{i, subject}) \quad (4)$$

where  $R_{i, subject}$  is the representation of the subject in the  $i^{th}$  trial, i.e., the features on subject noun.

The encoding-plus-retrieval model assumes that the representation of the subject changes due to feature percolation from the distractor as function of the percolation rate  $\theta$ . The representation of the subject thus comes from an encoding function with the percolation rate  $\theta$ :

$$R_{i, subject} \sim \text{Encoding}(\text{Input}_{subject}, \theta) \quad (5)$$

where  $Input_{subject}$  is the original input representation of the subject and  $\theta$  is the percolation rate. Using these equations, the encoding-plus-retrieval model allows us to make reading time predictions at the verb, which we now compare to reading time data from 11 experiments.

## Evaluation

To evaluate the encoding-plus-retrieval model, we first compare the prior predictions of the model against 11 interference effect datasets and then quantify the evidence for the model compared to the cue-based retrieval model for each dataset.

### Prior predictions

We generate prior predictions from the model assuming two free parameters, namely latency factor,  $F$  and percolation rate,  $\theta$ . The joint distribution of interference effect in grammatical and ungrammatical conditions,  $(E_{gram}, E_{ungram})$ , comes from encoding-plus-retrieval model conditional on  $F$  and  $\theta$

$$\begin{pmatrix} E_{gram} \\ E_{ungram} \end{pmatrix} \sim Model(F, \theta) \quad (6)$$

We specified the priors as follows. For the latency factor  $F$ , we used a truncated normal distribution:

$$F \sim Normal_{lb=0.1}(0.15, 0.03) \quad (7)$$

where  $lb = 0.1$  indicates a lower bound of 0.1 on latency factor values. Latency factor reflects average reading speed on a scale of 0 to 1. We choose this lower bound because a latency factor of less than 0.1 generates unreasonably fast reading times (see Jäger et al., 2017, for meta-analysis estimates of reading times). For the percolation parameter  $\theta$ , we chose a Beta distribution, which is bounded between 0 and 1 and has a peak near 0.3:

$$\theta \sim Beta(5, 10) \quad (8)$$

We have chosen a weakly informative prior on the percolation rate because we do not want to make any strong assumptions about this parameter.

The prior predictions of the model given the above priors are shown in figure 4. For each simulated trial, we sampled  $F$  and  $\theta$  from the priors and generated a retrieval time using the equations in the last section.

We compare the prior predictions of the model against the observed interference effect from 11 published datasets. If our model correctly captures the interference effect pattern in the grammatical and ungrammatical conditions, the mean observed effect from 11 datasets should lie within the 95% credible interval of the effect predicted by the model (Spiegelhalter, Freedman, & Parmar, 1994; Freedman, Lowe, & Macaskill, 1984). Figure 5(a) shows the 95% credible interval of the model’s prior prediction in the shaded gray area and the observed effects from each dataset.

The comparison of the model’s prior predictions against observed interference effect data shows that

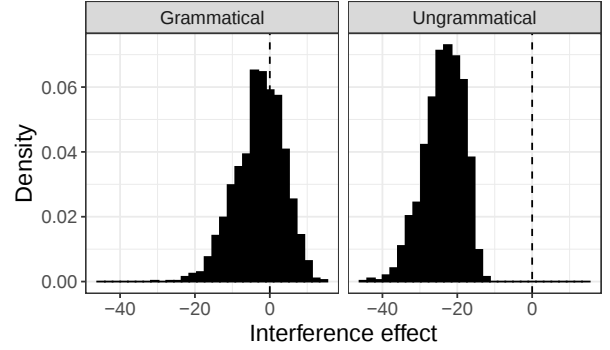


Figure 4: The prior predictive distributions for the grammatical and ungrammatical conditions generated by the encoding-plus-retrieval model.

1. The model correctly captures the pattern of interference effects in the grammatical conditions except for Dillon et al. (2013) experiment 1.
2. The model correctly predicts the facilitatory effect in the ungrammatical conditions. However, for a few datasets, the mean interference effect does not lie within the quantitative range predicted by the model; those are Dillon et al. (2013) experiment 1, Wagers et al. (2009) experiment 3 (singular verb), and Lago et al. (2015) experiment 3a and 3b.
3. The model generates reasonably constrained predictions. The predicted effect for grammatical conditions is in the range -16 to 8 milliseconds and for ungrammatical conditions in the range -35 to -15 milliseconds.

By contrast, the cue-based retrieval model cannot capture the observed range of effects in grammatical conditions as shown in figure 5(b). Thus, the prior predictions based evaluation indicate that the encoding-plus-retrieval model can better explain the range of interference effects compared to the cue-based retrieval model. We now compare the two models using Bayes factors.

### Model comparison

We quantify the evidence for encoding-plus-retrieval model compared to the cue-based retrieval model using Bayes factors. When comparing two models  $\mathcal{M}_1$   $\mathcal{M}_2$ , the Bayes factor is computed by taking the ratio of the marginal likelihoods given the two models. The Bayes factor,  $BF_{12}$ , is a real number value that represents the extent to which one model is more likely than the other, given the data. One widely followed convention in Bayesian modeling, which comes from Jeffreys (1939/1998), is to assume that a Bayes factor larger than 10 represents good evidence for the first model.

In order to estimate Bayes factors, we first estimated the marginal likelihoods for each model using Approximate Bayesian Computation (Sisson, Fan, & Beaumont, 2018; Palestro, Sederberg, Osth, Van Zandt, & Turner, 2018) and

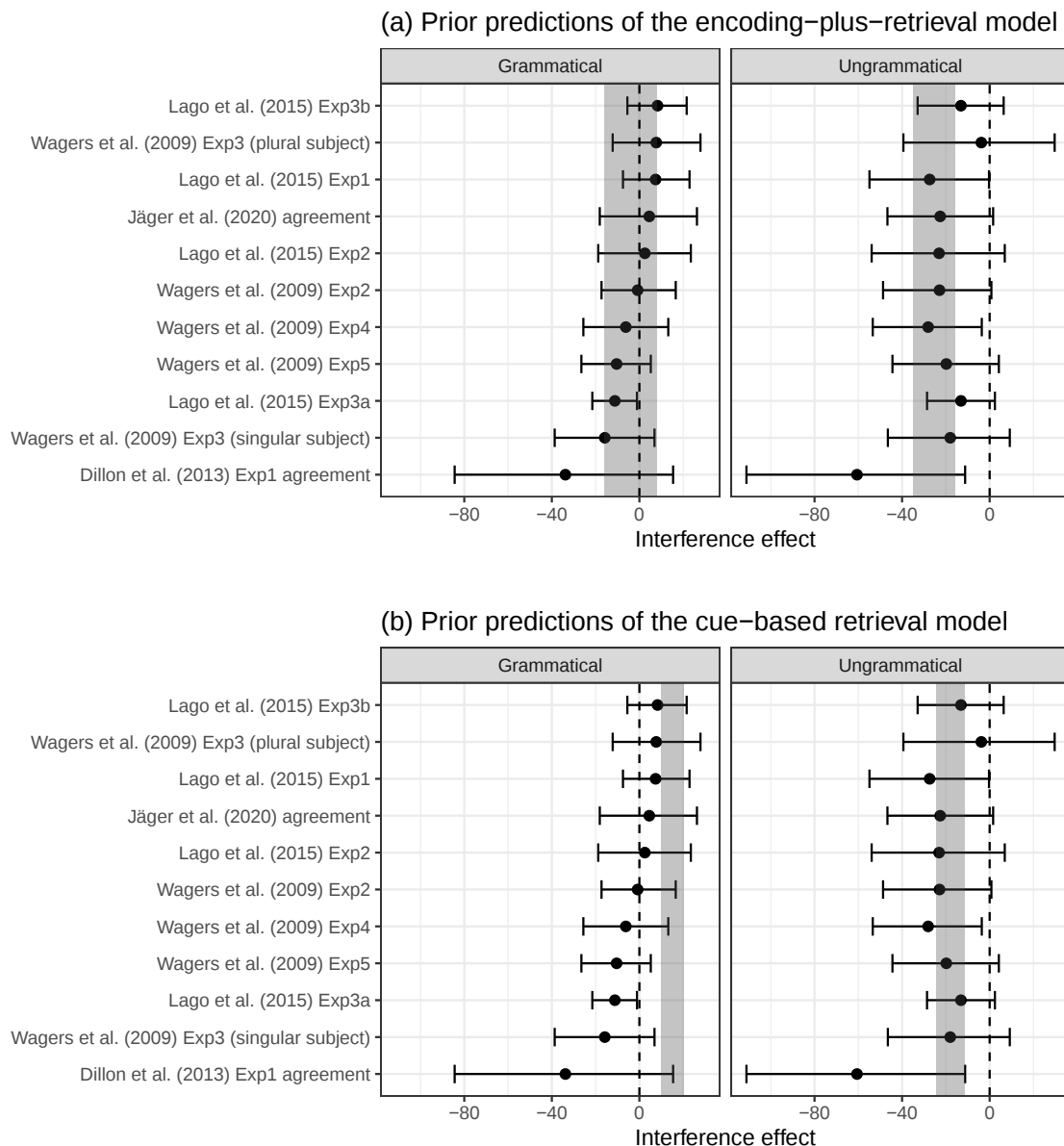


Figure 5: Comparison of the models’ predictions with the observed interference effect from 11 datasets. The shaded gray area represents the 95% credible interval of the prior predictions from the model.

then computed Bayes factors as the ratio of marginal likelihoods of the encoding-plus-retrieval model and cue-based retrieval model. Because Bayes factors can be very sensitive to the prior on the parameter of interest (here, percolation rate  $\theta$ ), we follow the suggestion in Schad et al. (2021) of choosing a range of priors on  $\theta$ , so that we can quantify evidence for the encoding-plus-retrieval model under different assumptions about the distribution of the percolation rate.

Figure 6 shows the estimated Bayes factor for each dataset given different priors on the percolation rate parameter,  $\theta$ . We find that Bayes factors are larger than 10 for all priors for 10 out of 11 datasets suggesting strong evidence in the favor of encoding-plus-retrieval model. Only Wagers et al.

(2009) experiment 3 (plural subject) data shows a better fit for the cue-based retrieval model. For all other 10 datasets, the encoding-plus-retrieval model clearly outperforms the cue-based retrieval model.

## Discussion

We have proposed a new model to account for the pattern of interference effects observed in grammatical and ungrammatical subject-verb agreement dependencies. The model assumes that dependency completion is driven by a cue-based retrieval process which is affected by a change in the representation of memory chunks due to feature percolation from a distractor noun. The evaluation of the model’s performance

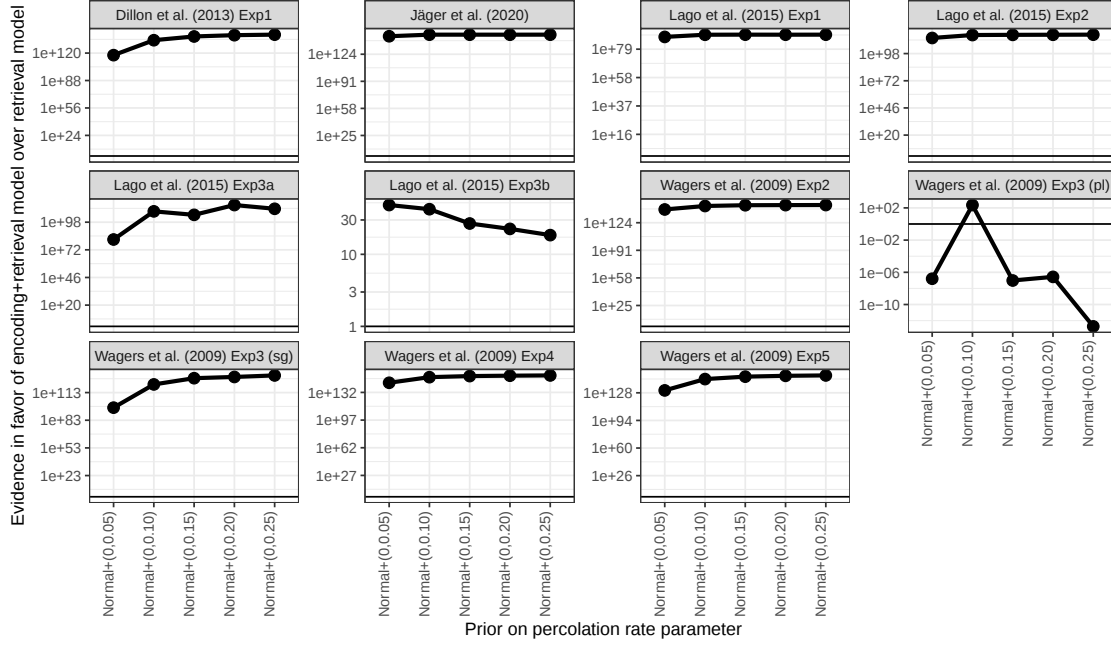


Figure 6: Estimated Bayes factors for 11 datasets given different priors on percolation rate parameter. Priors are shown on the X-axis. A prior of Normal+(0,0.05) means a zero-truncated normal distribution with mean 0 and standard deviation 0.05. The dashed line represents Bayes factor equal to one.

shows that

1. The prior predictions of the model correctly capture the observed pattern of interference effects in both grammatical and ungrammatical conditions.
2. The model outperforms the existing cue-based retrieval model of Lewis and Vasishth (2005) when fitted on the interference effect data from subject-verb agreement dependencies.

The results indicate that the cognitive processes underlying dependency completion may involve both encoding and retrieval mechanisms. This conclusion makes theoretical sense, because there is independent evidence for both types of mechanisms (Nairne, 1990; Lewis, Vasishth, & Van Dyke, 2006).

An interesting open question is whether the rate of feature percolation is sensitive to factors like the syntactic position of the distractor noun and syntactic distance between the distractor and the subject. It has been argued that there are constraints on how far a number feature can percolate in a tree (Nicol, Forster, & Veres, 1997; Eberhard, Cutting, & Bock, 2005). For instance, it could be more difficult for a feature to migrate from a noun in an embedded relative clause than from a noun in a prepositional phrase. Another constraint proposed for encoding-based accounts is that a feature can only percolate upwards in a syntactic tree (Eberhard, 1997; Vigliocco et al., 1995). Our model is currently agnostic to these factors. But they can be explored by developing constraints on feature percolation mechanism in our model. We plan to take this up in future work.

Another open question in the encoding-plus-retrieval model is the mapping between the percolation parameter  $\theta$  and participants' behavior. Because  $\theta$  is effectively the probability of misrepresentation of the subject, future work could estimate  $\theta$  from individual participants' error rates on comprehension questions that ask about the number marking on the subject. The individual-level  $\theta$  parameters estimated in this way could then be used to generate reading time predictions from the encoding-plus-retrieval model that could then be evaluated using the same participants' reading time data.

Note that we have implemented one variant of the encoding-based mechanisms in our model which assumes that in a certain trial the plural feature of the distractor can percolate to the subject noun with some probability. Another well-known variant is Marking and Morphing model (Eberhard et al., 2005) which assumes that the representation of the number feature of the subject noun is continuous and its distribution is disrupted by activation spread from the plural marking on the distractor noun. It would be interesting to compare encoding-plus-retrieval models under the above two assumptions about the encoding mechanism.

In sum, these modeling results reveal new insights about the constraints on sentence comprehension: encoding modulates retrieval. In other words, nouns stored in memory probabilistically undergo feature distortion, which in turn can affect how retrieval unfolds during dependency completion. To our knowledge, this is the first time that the joint impact of encoding and retrieval processes has been investigated computationally in sentence processing.

## Acknowledgments

We thank the three anonymous reviewers for helpful suggestions.

## References

- Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, 8(1), 57–99.
- Dillon, B. W., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69, 85–103.
- 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 syntax of sense: Number agreement in sentence production. *Psychological Review*, 112(3), 531.
- Engelmann, F., Jäger, L. A., & Vasishth, S. (2020). The effect of prominence and cue association in retrieval processes: A computational account. *Cognitive Science*. Retrieved from 10.1111/cogs.12800
- Freedman, L. S., Lowe, D., & Macaskill, P. (1984). Stopping rules for clinical trials incorporating clinical opinion. *Biometrics*, 40(3), 575–586.
- 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, 316–339. doi: 10.1016/j.jml.2017.01.004
- Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject–verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, 111. Retrieved from <https://psyarxiv.com/7c4gu/> doi: <https://doi.org/10.1016/j.jml.2019.104063>
- Jeffreys, H. (1939/1998). *The theory of probability*. Oxford University Press.
- Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement processes in Spanish comprehension. *Journal of Memory and Language*, 82, 133–149.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory and Cognition*, 18(3), 251–269.
- Nicol, J., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, 36, 569–587.
- Palestro, J. J., Sederberg, P. B., Osth, A. F., Van Zandt, T., & Turner, B. M. (2018). *Likelihood-free methods for cognitive science*. Springer.
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24(5 Series II), 574–590.
- Schad, D. J., Nicenboim, B., Bürkner, P.-C., Betancourt, M., & Vasishth, S. (2021). *Workflow techniques for the robust use of bayes factors*. (draft)
- Sisson, S. A., Fan, Y., & Beaumont, M. (2018). *Handbook of approximate bayesian computation*. CRC Press.
- Spiegelhalter, D. J., Freedman, L. S., & Parmar, M. K. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 357–416.
- 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.
- Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237.