The effect of prominence and cue association in retrieval processes: A computational account

Felix Engelmann
University of Manchester, UK

Lena A. Jäger
University of Potsdam, Germany

Shravan Vasishth
University of Potsdam, Germany

January 5, 2018

Abstract

We present a comprehensive empirical evaluation of the ACT-R based model of sentence processing developed by Lewis and Vasishth (2005) (LV05). The predictions of the model are compared with the results of a recent meta-analysis of published reading studies on retrieval interference in reflexive-/reciprocal-antecedent and subject-verb dependencies (Jäger, Engelmann, & Vasishth, 2017). We show that the model has only partial success in explaining the data, and that two independently motivated theoretical constructs are necessary: memory accessibility (prominence), and a theory of multi-associative cues. We implement these two constructs within the LV05 model and quantitatively compare the predictions of the original and the extended model with the results of the meta-analysis. Our simulations show that the extended model furnishes a superior fit. These results show that cue-based retrieval models need to take into account differences in the accessibility of items in memory, and the effect of context-based feature-selectivity. The simulations thus shed new light on the cognitive mechanisms underlying interference effects, and should be considered in the interpretation of existing empirical results and in the design of future experiments.

Keywords: ACT-R; Cue-based retrieval; Dependency completion; Retrieval interference; Computational modeling; Prominence; Multi-associative cues

Introduction

In psychology, memory access has long been argued to be a cue-based content-addressable mechanism (Anderson et al., 2004; Anderson & Lebiere, 1998; Ratcliff, 1978; Watkins & Watkins, 1975, among many others). These theoretical proposals have found

application in psycholinguistics, particularly in sentence comprehension research. One of these applications is the idea that the formation of non-adjacent linguistic dependencies relies on an associative cue-based retrieval process (Lewis, Vasishth, & Van Dyke, 2006; McElree, 2000; Van Dyke, 2002; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011).

Consider the linguistic processes that unfold at the verb phrase *was complaining* in Example 1. In order to understand who was doing the complaining, this verb phrase must be connected with a *noun phrase* that is *animate* and is a grammatical *subject* of the local clause where the verb phrase appears. These properties are assumed to be used as *retrieval cues* by an associative retrieval mechanism in order to seek out the correct linguistic dependent (here, *the resident*).

(1) The worker was surprised that **the resident** who was living near the dangerous neighbor **was complaining** about the investigation.

The retrieval processes activates the item in working memory whose features best *match* the retrieval cues. However, one of the core predictions of cue-based retrieval is that *similarity-based interference* arises between items in memory that simultaneously have features matching one or more of the retrieval cues. For instance, there are two other noun phrases in the example above that would match the *animate* cue: *the worker* and *the dangerous neighbor*. In addition, the noun phrase *the worker* is also a grammatical subject, although it is a subject of the main clause and not the local clause in which the verb phrase *was complaining* appears. When multiple noun phrases possess features that match one or more of the retrieval cues, this distracts attention from the correct noun phrase to be retrieved, affecting retrieval accuracy or retrieval time at the verb phrase *was complaining*.

Interference effects have been found to occur in other syntactic constructions as well. An example is the reflexive *himself/herself*. Consider Example 2 from Patil, Vasishth, and Lewis (2016).

(2) **The tough soldier** that Fred treated in the military hospital introduced **himself** to all the nurses.

Here, the reflexive *himself* requires a masculine-marked antecedent noun phrase to resolve its reference; this antecedent, *the tough soldier*, must be the subject of the main clause because English has a constraint that requires the antecedent to be in the same clause as the reflexive and in a particular syntactic relation with respect to the antecedent (called c-command, Reinhart, 1976). In this example, the constraint simply entails that the antecedent can only be the grammatical subject of the main clause. A noun phrase such as *Fred*, which appears inside the relative clause modifying the main clause grammatical subject, cannot be the reflexive's antecedent: it is in a *syntactically unlicensed* position. However, noun phrases in unlicensed positions could in principle cause interference at the reflexive if they possess a feature that is relevant for retrieval — in this case the masculine

Please send correspondence to felix.engelmann@manchester.ac.uk.

gender marking. The situation in reflexives is therefore similar to the case of subject-verb dependencies as shown in Example 1.

Numerous studies have found evidence for interference effects in subject-verb dependencies (Dillon, Mishler, Sloggett, & Phillips, 2013; King, Andrews, & Wagers, 2012; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Pearlmutter, Garnsey, & Bock, 1999; Tucker, Idrissi, & Almeida, 2015; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006, 2011; Wagers, Lau, & Phillips, 2009) as well as in reflexive-antecedent dependencies (Badecker & Straub, 2002; Chen, Jäger, & Vasishth, 2012; Cunnings & Felser, 2013; Felser, Sato, & Bertenshaw, 2009; Jäger, Engelmann, & Vasishth, 2015; Parker & Phillips, 2017; Sturt, 2003), although the situation in the case of reflexives is not without controversy (see, for example, Dillon et al., 2013).

One model that can explain such interference effects is the cue-based retrieval account of Lewis and Vasishth (2005), henceforth LV05. This model, which is based on the general cognitive architecture ACT-R ("Adaptive Control of Thought-Rational", Anderson et al., 2004; Anderson & Lebiere, 1998), is implemented as a parser that incrementally builds linguistic structure by carrying out a succession of memory retrievals to connect dependents such as subjects and verbs, and antecedents and reflexives. Based on the core assumptions of ACT-R that retrieving an item from memory is affected by activation decay and similarity-based interference, quantitative predictions for linguistic processing can be derived from the model and can be compared to empirical data. Over the last decade, the LV05 model has been widely used as a computational modeling framework by several research groups for investigating a range of empirical phenomena: (i) similarity-based interference effects (Dillon et al., 2013; Jäger et al., 2015; Kush & Phillips, 2014; Nicenboim, Logačev, Gattei, & Vasishth, 2016; Nicenboim & Vasishth, 2018; Nicenboim, Vasishth, Engelmann, & Suckow, 2018; Parker & Phillips, 2016, 2017; Patil, Vasishth, & Lewis, 2016; Vasishth, Bruessow, Lewis, & Drenhaus, 2008); (ii) the relative roles of predictive processing and memory effects (Boston, Hale, Vasishth, & Kliegl, 2011); (iii) impairments in individuals with aphasia (Mätzig, Vasishth, Engelmann, Caplan, & Burchert, 2018; Patil, Hanne, Burchert, De Bleser, & Vasishth, 2016); (iv) the interaction between oculomotor control and sentence comprehension (Engelmann, Vasishth, Engbert, & Kliegl, 2013); and (v) the effect of working memory capacity differences on underspecification ("good-enough" processing) in sentence comprehension (Engelmann, 2016; Vasishth & Engelmann, to appear).

Although the LV05 model has been applied to the study of specific theoretical questions, the empirical coverage of LV05 has never been quantitatively evaluated against a broad range of published findings. Such an evaluation is very important for at least two reasons. First, it serves as an important assessment of the model's capabilities and limitations. Modeling a single experimental result is informative but overfitting is an ever-present danger. Investigating multiple empirical results can yield a more realistic understanding of a model's performance, and understanding the range of the predictions that the model does (and does not) make is vital for evaluating model quality (Roberts & Pashler, 2000). Second, such a large-scale evaluation would allow other researchers to have a quantitative baseline for evaluating alternatives to the LV05 model. Recently, several alternative models to the LV05 parser have been proposed (Cho, Goldrick, & Smolensky, 2017; Rasmussen & Schuler, 2017; Smith, Franck, & Tabor, 2018), but no comprehensive model comparisons have been carried out against the full body of evidence available. Our large-scale evaluation

provides the foundation for such future work.

In this paper we derive the full range of predictions for interference effects of the LV05 model and compare them to the results of a recent meta-analysis by Jäger et al. (2017). In addition, we investigate two independently motivated principles and how they affect interference: *item prominence* and *multi-associative cues*.

Item prominence usually refers to the grammatical position of an item in the sentence or a certain discourse marking and has been acknowledged as an influential factor for the interference effect by several researchers (Cunnings & Felser, 2013; Patil, Vasishth, & Lewis, 2016; Van Dyke & McElree, 2011). The concept of multi-associative cues allows the associations between cues and features to be graded instead of categorical, accounting for the idea that the heuristics applied in memory retrieval processes are the result of context-dependent associative learning. We propose an implementation of item prominence and multi-associative cues and evaluate their effect on model predictions. We show that some patterns revealed by the meta-analysis cannot be explained by the LV05 model but can be accounted for by item prominence and multi-associative cues.

The paper is structured as follows. We begin by explaining the retrieval process assumed in the original LV05 model, and through simulation spell out the predictions for interference phenomena. We then discuss the empirical problems in the LV05 model by comparing its predictions to the Jäger et al. meta-analysis. Next, we introduce the constructs item prominence and multi-associative cues; we implement these concepts within the LV05 framework, and then present a series of simulations that evaluate the empirical coverage of the extended vs. original LV05 model. Compared to the original model, the inclusion of item prominence and multi-associative cues in the LV05 model yields an improved empirical coverage.

A comprehensive quantitative evaluation of the Lewis & Vasishth (2005) model

We first discuss the predictions of the LV05 model for interference effects in dependency resolution. As an empirical reference point, we use the Jäger et al. (2017) meta-analysis. We begin by describing the main type of constructions in this meta-analysis.

The Jäger, Engelmann, and Vasishth (2017) meta-analysis

The meta-analysis had data from 77 experimental comparisons from published eye-tracking and self-paced reading studies. Jäger et al. (2017) examined studies on subject-verb dependencies, reflexive-antecedent, and reciprocal-antecedent dependencies. We introduce the syntactic configurations that appeared in the meta-analysis, and also take this opportunity to introduce some terminology (in bold-face).

There were two classes of configuration in the meta-analysis. These are illustrated in Example 3 (Sturt, 2003). A retrieval is assumed to be initiated at the reflexive *himself* or *herself* in order to connect the reflexive with its antecedent. In all four sentences, the syntactically correct antecedent for the reflexive is the noun phrase *the surgeon*, whereas the other noun phrase *Jennifer* or *Jonathan* is inside a relative clause and thus not a syntactically legal antecedent of the reflexive (Chomsky, 1981). We therefore call the syntactically licensed antecedent the **target**, and the other noun phrase, which is in a syntactically

unlicensed position, the **distractor**.[1]

In all the sentences shown in Example 3, the grammatical gender of both target and distractor is manipulated. From a cue-based retrieval perspective, the distractor is assumed to interfere with the retrieval process whenever its gender matches the gender of the reflexive. In 3, the relevant retrieval cues and corresponding features are shown next to the reflexive and the two noun phrases, respectively. The relevant cues used for retrieval of the antecedent are *c-command*[2] and the gender of the reflexive *masculine* and *feminine*. There are other cues that could be used for retrieval but usually only two cues are relevant theoretically: One cue is used to differentiate between target and distractor (in the case of reflexives, c-command), and one cue is manipulated between conditions (in this case, gender). A + or − in front of the features of target and distractor indicates whether there is a **match** or a **mismatch** with the respective retrieval cue, which is shown on the reflexive in the examples below.

(3) a. *Target-match; distractor-mismatch*
The surgeon$^{\texttt{+MASC}}_{\texttt{+CCOM}}$ who treated Jennifer$^{\texttt{-MASC}}_{\texttt{-CCOM}}$ had pricked himself$\{^{\texttt{MASC}}_{\texttt{CCOM}}\}$...

b. *Target-match; distractor-match*
The surgeon$^{\texttt{+MASC}}_{\texttt{+CCOM}}$ who treated Jonathan$^{\texttt{+MASC}}_{\texttt{-CCOM}}$ had pricked himself$\{^{\texttt{MASC}}_{\texttt{CCOM}}\}$...

c. *Target-mismatch; distractor-mismatch*
The surgeon$^{\texttt{-FEM}}_{\texttt{+CCOM}}$ who treated Jonathan$^{\texttt{-FEM}}_{\texttt{-CCOM}}$ had pricked herself$\{^{\texttt{FEM}}_{\texttt{CCOM}}\}$...

d. *Target-mismatch; distractor-match*
The surgeon$^{\texttt{-FEM}}_{\texttt{+CCOM}}$ who treated Jennifer$^{\texttt{+FEM}}_{\texttt{-CCOM}}$ had pricked herself$\{^{\texttt{FEM}}_{\texttt{CCOM}}\}$...

In 3a and 3b, the target matches both cues `CCOM` and `MASC`, i.e., it is a **full match** for the reflexive. We will call these sentences **target-match configurations**. In 3c and 3d, the target does not match the gender of the reflexive and is thus only a **partial match** for the reflexive. Examples 3c and 3d will therefore be referred to as **target-mismatch configurations**. Note that, in this example, the gender match/mismatch of *the surgeon* only refers to its prototypical gender, which is masculine in English.

In 3b, the distractor *Jonathan* is a partial match for the reflexive because it matches the masculine cue. Under the content-addressable cue-based retrieval mechanism assumed in LV05, a partially cue-matching distractor is a potential retrieval candidate despite it being in a syntactically inaccessible position. Thus, the **distractor-match** condition 3b is assumed to induce retrieval interference in comparison with the **distractor-mismatch** condition 3a, where the distractor does not match the gender cue. The same distractor manipulation is applied in the target-mismatch configurations 3c and 3d.

In the LV05 model, there are two distinct types of interference effects expected in reading time data for target-match and target-mismatch configurations. The presence of a

[1] In the case of subject-verb dependencies such as Example 1, the target is differentiated from the distractor on the basis of it being the *local subject* for the verb while the distractor could be a subject but not the subject of the *local* phrase that contains the verb.

[2] Mostly for reasons of simplicity, c-command is usually represented as a static feature similar to gender, case, etc., although it is actually a syntactic *relation* between two items. It is therefore debatable whether some sort of syntactic search mechanism is needed to determine a c-command relationship or whether it is approximated in some other way, e.g., by a *subject* and a *local-clause* feature. See Kush (2013) for an investigation of the computational complexity needed for keeping track of c-commanders.

partially matching distractor might either slow down or speed up reading at the critical region, i.e., at the reflexive, the reciprocal, or the verb depending on the syntactic construction being considered. Slow-downs and speed-ups are interpreted as **inhibitory interference** and **facilitatory interference**, respectively, meaning that the presence of a distractor leads to an **inhibition** or a **facilitation** during the retrieval process. As we explain below, in the LV05 model, inhibitory effects are expected in target-match configurations, whereas facilitatory effects are expected in target-mismatch configurations.

Predictions of the Lewis & Vasishth (2005) model for target-match and target-mismatch conditions

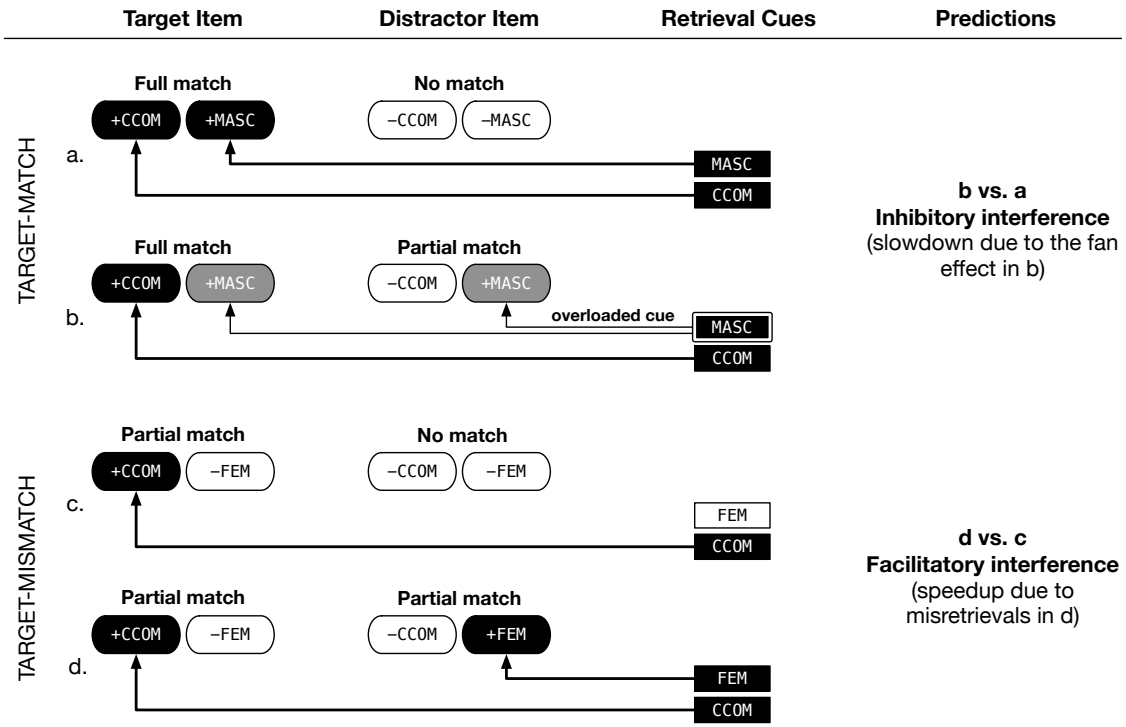


Figure 1. Predictions of ACT-R for the four conditions shown in Example 3. Line weights indicate the amount of spreading activation from a cue to an item. Black oval boxes represent a feature match. Gray oval boxes indicate features matching an 'overloaded' cue (`MASC` in b), and white boxes indicate a mismatch.

See Figure 1 for a graphical representation of the model predictions for Example 3. The oval boxes indicate matching (black or gray) or mismatching (white) features of an item with respect to the retrieval cues. The darker the boxes the better is the match of the item and the higher is its activation level. The relative activation levels of memory items in ACT-R determine their retrieval accuracy and retrieval speed. The item with the highest activation at the time of retrieval has the highest probability of being retrieved and the fastest retrieval time. Each item has a *base-level* activation that reflects past usage by accounting for all reactivations (i.e., access events at times t_j) and a time-based decay with

rate d (this usually has the default value 0.5 in ACT-R):

$$B_i = \ln(\sum_{j=1}^{n} t_j^{-d}) + \beta_i \tag{1}$$

In addition to the base-level, *spreading activation* is added to every (partially) matching item at the time of retrieval. The spreading activation component is the main source of similarity-based interference effects in ACT-R. An item receives spreading activation from all matching cues j depending on the *associative strength* S_{ji} between cue j and item i and the cue's weight W_j. W_j is standardly set to one divided by the number of cues, meaning that all cues are weighted equally. We are adopting this standard assumption throughout this work.

$$S_i = \sum_j W_j S_{ji} \tag{2}$$

The arrows in Figure 1 show how activation from the retrieval cues is distributed to the target and the distractor based on their features. The thickness of the lines with arrows indicates the amount of spreading activation that is added to an item due to that feature, assuming that each cue is weighted equally. In Figure 1a (cf. Example 3a), the target is a full match for the set of retrieval cues, `MASC` and `CCOM`. Both cues are also **unambiguous** because they are matched by the target only and not by the distractor. The target thus receives the maximal amount of spreading activation at retrieval. In the interference condition b in Figure 1 and Example 3, in contrast, the gender cue is matched by the distractor in addition to the target. Thus, the `MASC` cue is now **ambiguous** or "overloaded" (Watkins & Watkins, 1975), with the result that the activation from this cue is now split between the target and the distractor. This follows from Equation 3: The associative strength between a cue and an item is reduced in relation to the **fan** — the number of items associated with the cue (MAS is the value of the *maximum associative strength*).

$$S_{ji} = MAS - \ln(fan_j) \tag{3}$$

Each cue distributes the *limited* available activation equally between all matching items (with the maximally available amount being $W_j \times MAS$). The more competitor items are present that match a cue j, the weaker the association of this cue with an item i. Each competitor thus takes away some amount of spreading activation from the target item and thus makes it harder to distinguish from the other items. This is called the **fan effect** (Anderson, 1974). In our example (Figure 1 and Example 3), the fan effect causes a reduction of the spreading activation received by the target in b in comparison with a, thus reducing the target's total activation, which is the sum of the base-level B_i and the spreading activation S_i plus random noise ϵ_i (cf. Equation 4). A decrease in activation causes retrieval time RT_i to increase. As shown in Equation 5, retrieval time is a negative exponential function of the total activation at the time of retrieval, where F and f are two scaling parameters — the *latency factor* and the *latency exponent*, respectively.

$$A_i = B_i + S_i + \epsilon_i \quad (4)$$

$$RT_i = Fe^{-(f \times A_i)} \quad (5)$$

Hence, the similarity in gender between target and distractor in target-match configurations shown in Figure 1 a vs. b predicts a slower retrieval latency due to the fan effect. We will refer to this slow-down as an **inhibitory interference effect**. In addition, there is a higher probability in b compared to a that the distractor is erroneously retrieved instead of the target. This is because activation in ACT-R fluctuates due to the noise component in Equation 4. We refer to retrievals of the distractor as **misretrievals**.[3]

The predictions for retrieval time are different in target-mismatch configurations c and d of Figure 1 and Example 3. In c and d, the target is only a partial match as it does not exhibit the correct gender feature `+FEM`. When the distractor matches the gender in d, there is, however, no reduction in the target's activation. The reason is that both cues `FEM` and `CCOM` are only matched by one item each and are thus not ambiguous. Hence, no fan effect and no inhibitory interference is predicted. However, since target and distractor now both receive the same amount of spreading activation — each matches exactly one cue — their activation levels are relatively close to each other. As a consequence, both items enter into a *race process* where the retrieval of either item is almost equally probable. A race process has the effect that, on average, the retrieval latency is shorter than when there is a clear winner due to a bigger difference in activation as is the case in condition c (e.g., Van Gompel, Pickering, & Traxler, 2001; for simulations demonstrating a race process, see Logačev & Vasishth, 2016). Hence, the prediction for target-mismatch configurations in Figure 2d vs. 2c is a speed-up. We refer to this speed-up as a **facilitatory interference effect**.

Comparison of the LV05 predictions with the results of the meta-analysis

Figure 2 summarizes the predictions of ACT-R for interference effects for simulations of target-match and target-mismatch configurations.[4] The figure shows the possible ACT-R predictions for interference effects on retrieval latency over a range of values for the most relevant parameters. The interference effect is calculated as the latency difference between distractor-match and distractor-mismatch conditions (distractor-match − distractor-mismatch) *within* target-match and target-mismatch configurations, so that values above zero indicate *inhibitory* interference (slow-down) and values below zero indicate a *facilitatory* effect (speed-up). Along the x-axis of Figure 2, we plot increasing values of the latency factor F, which is usually the most freely varied parameter in ACT-R models and simply scales the retrieval latency. While there is variation in the mean interference effect along

[3] Note that in an alternative model of cue-based retrieval proposed by McElree, Foraker, and Dyer (2003), the direct-access model, interference is only reflected in a decreased retrieval probability of the target but not in retrieval time. Effects observed in reading times are then explained as a by-product of changes in the retrieval probabilities. The idea here is that misretrievals may trigger a repair process that inflates reading times (McElree, 1993). For an implementation and quantitative comparison of the direct-access model with the LK05 model, see Nicenboim & Vasishth, 2018.

[4] Simulations were carried out in R (R Core Team, 2016). The code is available at https://github.com/felixengelmann/inter-act.

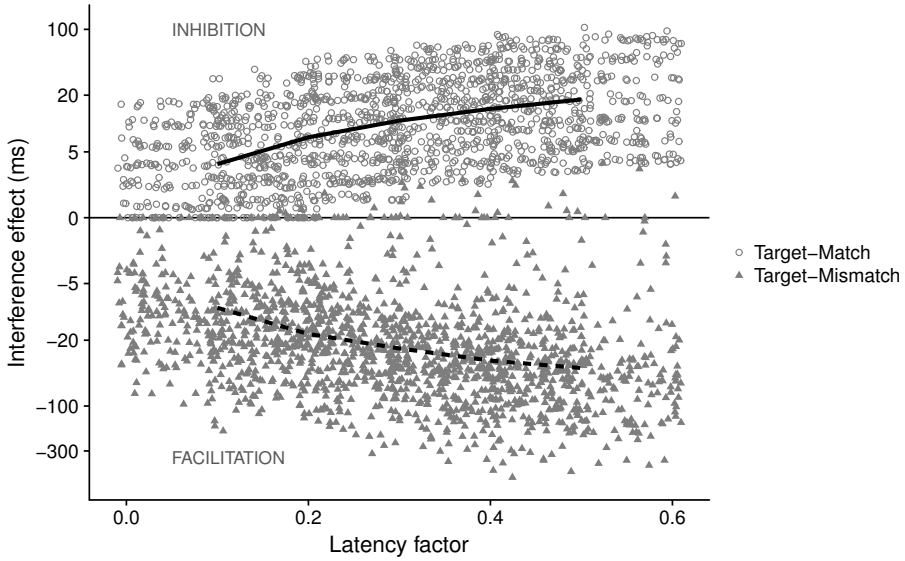


Figure 2. ACT-R predictions for the interference effect (distractor-match − distractor-mismatch) in target-match and target-mismatch configurations as a function of the latency factor, and for a range of parameter values (768 combinations): Latency factor $0.1 \leq F \leq 0.5$, noise $0.1 \leq ANS \leq 0.3$, maximum associative strength $1 \leq MAS \leq 4$, mismatch penalty $0 \leq MP \leq 2$, retrieval threshold $-1 \leq \theta \leq 0$. Lines represent the mean effect. Y-axis is log-scaled.

different parameter values, the figure clearly shows that the predictions of the LV05 model are restricted to *inhibitory interference* in *target-match* configurations (caused by the fan effect) and *facilitatory interference* in *target-mismatch* configurations (caused by the race process between target and distractor).

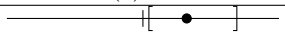
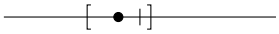
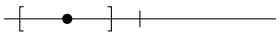
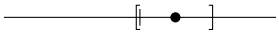
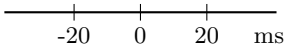
How well do these predictions fare compared to the evidence published in the literature? It turns out that the answer is: not very well. Table 1 summarizes the quantitative results of the meta-analysis for the interference effect showing the mean posterior effect estimates and the posterior probability $P(b > 0)$ of the effect being greater than 0.[5] In Table 1, we show the effects of interference in target-match and target-mismatch configurations for each of the dependency types separately.

In Jäger et al. (2017), subject-verb dependencies were divided into *agreement* dependencies (e.g., Pearlmutter et al., 1999; Wagers et al., 2009) and *non-agreement* dependencies (e.g., Van Dyke, 2007; Van Dyke & McElree, 2011), because these constitute two distinct lines of research, usually showing different patterns. While agreement studies have focused on effects of number attraction, non-agreement studies investigated interference effects in-

[5]Note that the posterior probability should not be confused with the frequentist p-value; the posterior probability here is the probability of the effect being positive, and therefore there is no concept of a conventional cut-off critical value such as 0.05, and a binary decision of "significant" or "non-significant" would be misleading.

Table 1

Results of the Jäger et al. (2017) Bayesian meta-analysis showing mean posterior effect estimates $\bar{b}$ with Bayesian 95% credible intervals in the Evidence column and the posterior probability of the effect being greater than 0. The credible interval represents the range over which we can be 95% certain that the true value of the estimated effect lies, given the data (note that the posterior probability should not be confused with the frequentist p-value). A positive interference effect means inhibition, a negative one facilitation. Results are compared with the predictions of cue-based retrieval as implemented in the LV05 ACT-R model and the additional contributions of item prominence (IP) and multi-associative cues (MAC).

Dependency	Target	Evidence ($\bar{b}$)	$Prob(b > 0)$	ACT-R	+IP	+MAC
Subject-verb non-agreement	Match		0.99	✓		
Subject-verb agreement	Match		0.09	✗	✓	
	Mismatch		0	✓		
Reflexives/ Reciprocals	Match		0.53	✗	✓	
	Mismatch		0.97	✗		✓
						

volving other semantic and syntactic cues. Reflexive-antecedent and reciprocal-antecedent dependencies were treated as one category in the meta-analysis because both follow the same syntactic constraint and the data of only two publications on reciprocals were available when the Jäger et al. (2017) article was published.

Clearly, the model cannot account for all the findings of the meta-analysis shown in Table 1. In *target-match* configurations, the predicted inhibitory effect was found only for non-agreement subject-verb dependencies. The other dependency types did not provide enough evidence for any effect in target-match configurations; however, these cases may not necessarily be problematic for the model because of the generally low power of the published studies (see Jäger et al., 2017; Nicenboim et al., 2018 for discussion). Most problematic for the model predictions in target-match configurations are individual studies that found a facilitatory effect. For *target-mismatch* configurations, the prediction of a facilitatory effect is only supported by subject-verb agreement studies; reflexive-/reciprocal-antecedent dependencies show inhibition. For non-agreement subject-verb dependencies, no target-mismatch data were available at the time of the meta-analysis. However, two recent studies show evidence for the predicted facilitatory effect in target-mismatch configurations in reflexives (Parker & Phillips, 2017) and in non-agreement subject-verb dependencies (Cunnings & Sturt, 2018).

As discussed in Jäger et al. (2017), one important observation here is that in both target-match and mismatch configurations, the individual results of different studies show a considerable range of variability, ranging from facilitatory to inhibitory interference. A closer look at the experimental designs of individual studies sheds more light on some of the reasons for this variability. Table 1 also indicates in columns six and seven that some of the unexplained facilitatory effects in target-match configurations can be explained in terms of item prominence (IP), and inhibitory effects in target-mismatch configurations can be explained by a theory of multi-associative cues (MAC). We discuss item prominence and multi-associative cues in the next two sections, followed by their formal implementation in the model.

Item prominence

With item prominence, we refer to a relation of a linguistic item to other items within a sentence or discourse context, that makes this item particularly *salient* in memory. This relation can be a distinguishing grammatical position in the sentence (e.g., subject vs. object) or a discourse marking such as topicalisation. Both qualities have been discussed in the literature (Cunnings & Felser, 2013; Patil, Vasishth, & Lewis, 2016; Sturt, 2003; Van Dyke & McElree, 2011). The model we present in this paper integrates item prominence into ACT-R. The model can thus be used to predict the consequences that different prominence levels have on the interference effect under the general assumptions of ACT-R.

Independent evidence shows that the accessibility of a noun phrase is increased in prominent grammatical positions or through increased discourse saliency, such as topicalization (Ariel, 1990; D. Arnold, 2007; Brennan, 1995; Chafe, 1976; Du Bois, 2003; Grosz, Weinstein, & Joshi, 1995; Keenan & Comrie, 1977). Thus, for our model, we assume that the prominence of an item affects its general activation in memory, independent of how well its features match the retrieval cues. In other words, a prominent item is more *salient* or more *accessible* in memory than a low-prominence item and this should influence the interference effect during retrieval. Thus, a sentence containing a high-prominence distractor should show a different interference effect than a sentence with a low-prominence distractor, even if the target and the retrieval cues are the same. Expressed in ACT-R terms, a high prominence status results in an increased *base-level activation* B_i, which is the activation of an item before *spreading activation* S_i is added as the result of the retrieval cues.

Figure 3 shows the predictions of our model as a function of the prominence of the distractor (in terms of its base-level activation) with respect to the prominence of the target. The x-axis represents the difference in base-level activation between target and distractor while the target activation stays constant at 0. The y-axis shows the predicted interference effect in target-match and target-mismatch configurations.

We first look at the predictions for *target-mismatch* configurations (broken line) as these are more straightforward than the target-match predictions. Recall that the interference effect in target-mismatch configurations is caused solely by a race process between two similarly activated items. Because no features are overlapping between the items, no fan effect is predicted and, hence, no inhibition. The facilitatory effect increases with distractor prominence, reaching its maximum when the distractor activation is equal to the target activation (as discussed in Logačev & Vasishth, 2016, facilitation in a race process is largest when the two racing processes have similar completion times, which would be the

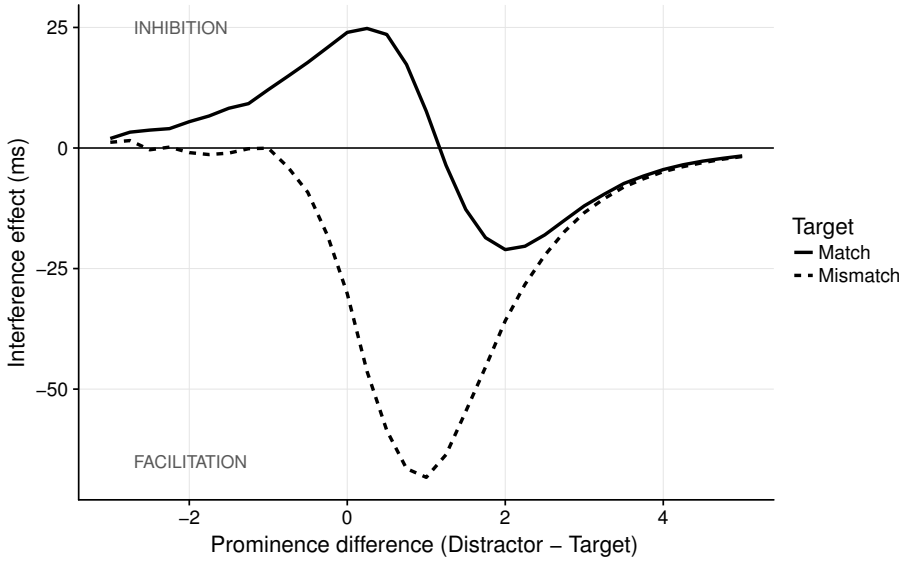


Figure 3. Predicted target-match and target-mismatch interference effects as a function of the difference between distractor activation and target activation (when target activation is constant).

case when the distractor and target have very similar activation values). For higher activation differences, the target-mismatch facilitation effect decreases again when the distractor activation exceeds the target activation such that interference from the target eventually becomes negligible and only the distractor is ever retrieved.

In *target-match* configurations (solid line in Figure 3), we see two major predictions: the inhibitory target-match interference effect (A) *increases* with increasing distractor activation, and (B) *decreases* when the distractor activation exceeds the target activation, eventually turning into a *facilitatory interference effect*. This facilitation is explained by a race process that occurs between similarly activated retrieval candidates just as is the case in target-mismatch configurations. This race-based facilitatory effect counteracts the inhibitory fan effect in target-match configurations when the distractor is very prominent, i.e., highly activated, compared to the target. For the remainder of this section, we concentrate on the predictions in target-match configurations.

In the literature on target-match interference configurations with high-prominence distractors, there is some evidence for both (A) increasing inhibitory effects as well as (B) facilitatory effects.

A: Increasing target-match inhibition. In an eyetracking and SAT experiment with target-match configurations, Van Dyke and McElree (2011) found that a distractor noun phrase in the subject position of a subordinate clause, such as *the witness* (vs. *motion*) in 4a, causes inhibitory interference at the main verb *compromised*, while no such effect was present when the distractor *the witness* was in object position as in 4b.

(4) a. The judge who had declared that **the witness/the motion** was inappropriate

realized that the attorney in the case compromised.

b. The judge who had rejected **the witness/the motion** realized that the attorney in the case compromised.

Patil, Vasishth, and Lewis (2016) found an interference effect at the reflexive in an eyetracking experiment using sentences as in 5 with the distractor *Fred* in subject position, which was a manipulation of Sturt (2003) in 6 where the distractor was in object position.

(5) The tough soldier that **Fred**/Katie treated in the military hospital introduced himself to all the nurses.

(6) The surgeon who treated **Jonathan**/Jennifer had pricked himself with a used syringe needle.

In both the manipulations of Van Dyke and McElree (2011) and Patil, Vasishth, and Lewis (2016), a prominent distractor (in subject position) in a target-match configuration caused inhibitory interference while a non-prominent distractor (in object position) did not. A strengthened interference like this as a consequence of higher prominence is predicted by our prominence model as shown in Figure 3.

In a reflexive-antecedent study in Chinese Mandarin, Jäger et al. (2015) found a similar difference in target-match configurations between Experiment 1, where a distractor was present in the sentence, and Experiment 2, where three distractors were presented as memory load. An inhibitory target-match interference effect was only found in Experiment 2. In addition to the higher number of distractors in Experiment 2, the need to rehearse the distractors throughout sentence comprehension would make them more prominent in memory, i.e., increase their activation, which would amplify the interference effect, again as predicted in Figure 3.

B: Target-match facilitation. Sturt (2003, Exp. 1) and Cunnings and Felser (2013, Exp. 2) found *facilitatory interference in target-match configurations* when the distractor was in subject position *and* made the discourse topic using a context sentence. Cunnings and Felser used sentences such as Example 7, where the distractor noun phrase was introduced in a context sentence and was referred to in the target sentence through the pronoun *he*. In their discussion section (pp. 212–213), Cunnings and Felser hypothesized that the distractor was more prominently encoded due to reactivation at the anaphora, and that this may have increased the probability of discovering an interference effect at the reflexive.

(7) **James** has worked at the army hospital for years.
The soldier that **he** treated on the ward wounded himself while on duty in the Far East.

A topicalized distractor that is also in subject position would arguably be more prominent than if it were just in subject position but not topicalized. The qualitative difference of the target-match effects described above is thus predicted by our model. As Figure 3 shows, these patterns are the consequence of (A) an increased fan effect with increased distractor prominence, and (B) a facilitatory effect through an emerging race between target and distractor for very high distractor prominence.

In summary, the integration of prominence in the form of memory activation can explain findings of *inhibitory* interference in target-match configurations with a prominent distractor that were not found with a non-prominent distractor (Jäger et al., 2015; Patil, Vasishth, & Lewis, 2016; Van Dyke & McElree, 2011), and findings of *facilitatory* interference in target-match configurations with a highly prominent distractor that was in subject position *and* the discourse topic (Cunnings & Felser, 2013; Sturt, 2003). The original LV05 model neither predicts facilitatory interference effects in target-match configurations nor the systematic absence of an effect under certain conditions. Earlier, in Table 1, we had shown the explanatory gaps of LV05 with respect to the outcomes of the Jäger et al. (2017) meta-analysis, specifically the facilitatory interference effect in target-match configurations in subject-verb agreement and the absence of an overall effect in reflexives and reciprocals. Taking into account item prominence as presented above, these unexplained effects are possible outcomes on the continuum shown in Figure 3.

Also contained in Table 1 is the finding of *inhibitory* instead of facilitatory interference in *target-mismatch* configurations. This has been found, for instance, in some studies on reflexives and reciprocals and can neither be explained by LV05 nor by item prominence. According to ACT-R, inhibitory interference simply cannot arise in target-mismatch configurations because the necessary condition for a fan effect — an overloaded cue due to multiple matches — is not met. The next section will introduce the concept of *multi-associative cues*, which proposes a less categorical cue-feature matching that predicts inhibitory interference for certain highly constrained retrieval contexts even in target-mismatch configurations.

Multi-associative Cues

Conceptualizations of cue-feature relations are often simplified as being categorical in at least two ways. One simplification mostly made is that a match between a retrieval cue and feature is assumed to be binary: Whether an item matches, e.g., the `masculine gender` cue is a yes-or-no decision. Another assumption made by models such as ACT-R is that the matching is slot-based: A gender cue can only be matched by gender features (`masc`, `fem`, `neut`) and is not associated at all with features of a different category. With the proposal of multi-associative cues, we argue for a more abstract and graded view of retrieval cues. The proposal is motivated by a usage-based approach to language acquisition, which views the cognitive representation of language as the representation of one's individual experience with form, meaning, and context (Bybee, 2006; Langacker, 1987; Tomasello, 2003). Rule-like behavior in language processing is therefore not based on clear-cut categories but emerges from analogy between similar form-to-function relations. In the same manner, our notion of multi-associative cues is that retrieval cues represent abstract knowledge about the properties that successfully identify the correct retrieval target, as derived from experience with a certain dependency context. Hence, cue-feature relations evolve as graded associations between a retrieval context and any properties of the correct target resulting from a process of learning relevant discriminations between features.

As a result, a cue can be associated with *multiple* feature values at varying degrees. A cue might also be equally associated with two different features in certain contexts where either of these two features can be used to successfully identify the target.

Expressed in terms of classical conditioning, two stimuli (features) are discriminated when they elicit different responses (Rescorla & Wagner, 1972). Two stimuli that require

similar responses in similar context will be less discriminated. For the case of linguistic dependency resolution, this would mean that two features that frequently co-occur in a retrieval context (e.g., the same type of dependency) in identifying a target will be less discriminated than two features that also occur in combinations with other features.

As an example, consider the difference between the retrieval contexts of reflexives and of reciprocals, shown in Table 2. The correct antecedent for an English reflexive can exhibit different feature combinations depending on the specific form of the reflexive, i.e., *himself*, *herself*, *itself*, and *themselves*. A dissociation of c-command, gender, and also number in the retrieval request is therefore necessary for identifying the correct target with respect to the individual form of the reflexive. Therefore, the retrieval cues in English reflexives are *highly selective*.

In contrast, correct targets for a reciprocal invariably exhibit the features `+PLUR` and `+CCOM`. Because of their invariable co-occurrence, an effective retrieval cue specification for reciprocals does not require a strong discrimination between `+PLURAL` and `+CCOM`. Instead, it can be thought of as more efficient to also activate plural items with the `CCOM` cue and vice versa. This is a case of *low selectivity*. As a result, in the context of a reciprocal-antecedent dependency, the cues `CCOM` and `PLURAL` would both be associated to some degree with both the features `+CCOM` and `+PLUR`, i.e., they are *cross-associated*. A similar situation arises for the Chinese reflexive *ziji* (also shown in Table 2), which requires an animate c-commanding target. Thus, in the case of *ziji*, `CCOM` would be cross-associated with `ANIM`.

Table 2
Possible feature combinations exhibited by correct antecedents of English reflexives, reciprocals, and Chinese ziji.

Context	Target features	Form
EN reflexive	{+MASC, +CCOM}	himself
	{+FEM, +CCOM}	herself
	{+NEUT, +CCOM}	itself
	{+PLUR, +CCOM}	themselves
EN reciprocal	{+PLUR, +CCOM}	each other
CN reflexive	{+ANIM, +CCOM}	ziji

Figure 4 illustrates a case of cross-association in a reciprocal-antecedent dependency in *target-mismatch* configurations in our extended ACT-R model. Because the `CCOM` and `PLUR` cues are cross-associated, both cues behave here as a kind of amalgamated cue that is associated with both the `+CCOM` and the `+PLUR` feature. In the target-mismatch/distractor-mismatch condition c, the target therefore receives activation from both cues although it only carries the `+CCOM` feature. In the target-mismatch/distractor-match condition d, the target carries `+CCOM` and the distractor carries `+PLUR`. As a consequence, both cues now share their activation between target and distractor, i.e., they are overloaded. This leads to a similar situation as in target-match configurations shown earlier in condition b of Figure 1: As spreading activation is shared between target and distractor, inhibitory interference, i.e., a *fan effect*, arises. This is because both items are less activated in d than the target is in

c and will be retrieved slower in d vs. c.

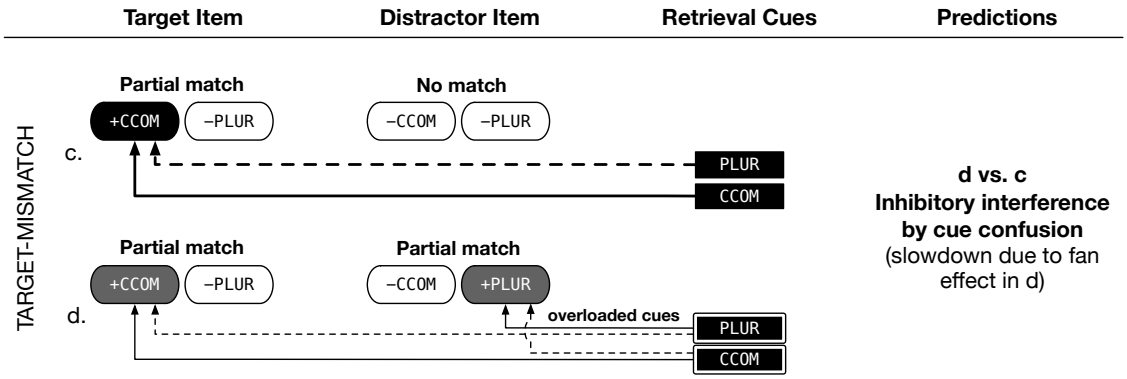


Figure 4. Predictions of the extended ACT-R model showing the consequences of cross-association in target-mismatch configurations of reciprocals. Line weight and box shading indicate the amount of spreading activation added to an item due to a feature match. Dashed lines represent spreading activation to a cross-associated feature.

Figure 5 shows the predictions of an increasing *cross-association level* in our model. In target-mismatch configurations, a higher cross-association causes an inhibitory fan effect that eliminates the facilitatory effect.

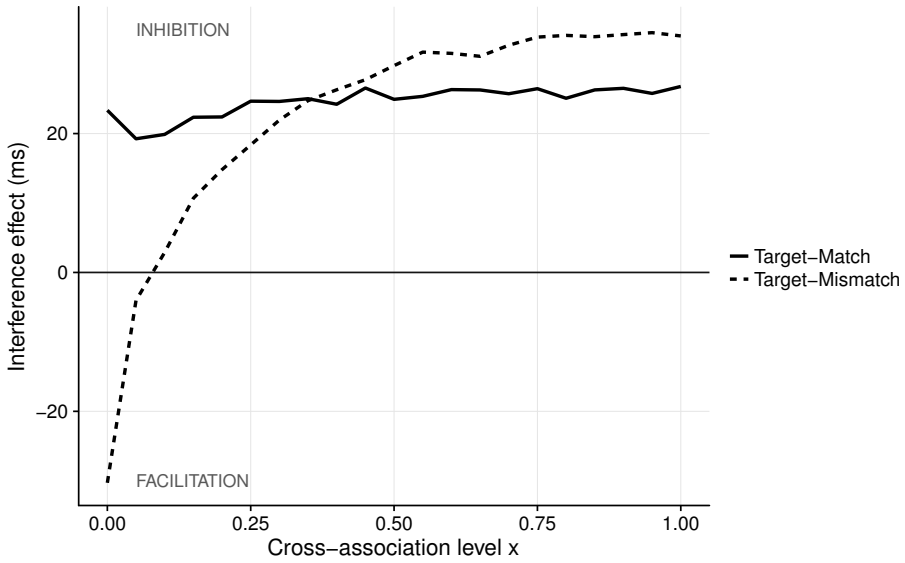


Figure 5. Predicted interference effect as a function of the cross-association level x.

The cross-association level c takes values between 0 and 1, where $c = 0$ means that two features are maximally discriminated (distinct cues activate distinct features) and $x = 1$ means that their corresponding features are treated as functionally identical, i.e., each cue activates both features.

More formally, $C_{kl}(Context)$ is the cross-association level x with respect to features k and l in a particular retrieval context (e.g., English reciprocals), and is equal to the strength

with which each feature is associated with the corresponding cue of the other feature. For example, if the cross-association level of `+CCOM` and `+PLURAL` in reciprocals equals 0.5, it means that the `CCOM` cue is associated with the `+PLURAL` feature with strength 0.5 and the `PLURAL` cue is associated with the `+CCOM` feature with strength 0.5. This means that, in the absence of the plural cue, a plural item would still receive activation from the cue `CCOM`, but the plural item would not receive as much activation as a c-commanding item would. Thus, at $c = 0.5$, there is still some discrimination between the features in question. If, however, $c = 1.0$, plural and c-command would not be discriminated at all as distinct information. Any item with one of the two features would be activated by any of the two cues in the same way. This effectively means that we would not think of two cues in this case but only one that is associated equally with two features.

Theoretically, the cross-association level c reflects the relative frequency of co-occurrence of both features, relative to the frequency of occurrence of either of the features. For example, consider Table 2, which shows several co-occurring features. We could say that the cross-association level $C_{kl}(Context) = c$ is the ratio of all feature combinations with both k and l with respect to all combinations with at least k or l, given a particular context:

$$C_{kl}(Context) = \frac{\sum [k \wedge l | Context]}{\sum [k \vee l | Context]}, \tag{6}$$

where the square brackets represent an Iverson bracket which denotes 1 if the enclosed condition is satisfied and 0 if not. This way, we can say, e.g., that the cross-association levels for the examples in Table 2 are for reflexives $C_{\texttt{CCOM,MASC}}(\textit{refl-EN}) = 1/4 = 0.25$, for reciprocals $C_{\texttt{CCOM,PLUR}}(\textit{reci-EN}) = 1/1 = 1.0$, and for *ziji* $C_{\texttt{CCOM,PLUR}}(\textit{ziji}) = 1/1 = 1.0$. The absolute values of these parameters are not of importance here; this example only serves as an illustration of the difference between English reflexives on the one hand and reciprocals or *ziji* on the other. What this calculation suggests is that, in processing English reflexives, more distinct cue representations are used due to a greater variety of feature combinations than in reciprocals or *ziji*.

In summary, the theory of multi-associative cues predicts that a cue would in some situations share its spreading activation between what would otherwise be categorically distinct features. In these situations, a fan effect can arise even in *target-mismatch* configurations and not only in target-match configurations as in standard ACT-R. The theory predicts a higher cross-association level for both reciprocals and the Chinese reflexive *ziji* compared to English reflexives. This could explain the result of Kush and Phillips (2014), who found inhibitory interference in target-mismatch conditions in Hindi reciprocals, as well as our finding of an inhibitory target-mismatch effect for Chinese *ziji* in Experiment 1 of Jäger et al. (2015). The following section explains the implementation of both multi-associative cues and item prominence in our extended ACT-R model.

Implementation of item prominence and multi-associative cues

The ACT-R architecture already has the basic theoretical constructs needed for implementing prominence and multi-associative cues. For example, in ACT-R, any two memory items can be assigned a numerical value that signifies how similar they are to each other. Thus, the colors orange and red can be treated as more similar to each other than orange and green. Because feature values are also treated as items in memory, similarities can be

assigned to pairs of features as well. In ACT-R, similarities are used, for example, in the equation for a component called *mismatch penalty* that enables the model to retrieve items that do not match the retrieval cues but might nevertheless be similar. Thus, an orange item can be retrieved even though the retrieval cue specifies a red one. We extend the ACT-R framework so that the similarity between features is also used in the computation of the fan effect.

In order to incorporate a mechanism for prominence and multi-associative cues, we redefine the associative strength S_{ji} between a cue j and an item i. Recall from Equation 4 that, given a set of retrieval cues ($Cues = \{c1, \ldots, cJ\}$), the activation A_i of an item i is partly a function of spreading activation S_i:

$$A_i \propto S_i \text{ where } S_i = \sum_{j \in Cues} W_j S_{ji} \tag{7}$$

For each cue j, the standard ACT-R calculation of S_{ji} is based on the fan, which is defined as the number of items that match this cue. Instead of this simplified definition, we base our implementation on the more general definition of S_{ji} (Schneider & Anderson, 2012, p. 129). This general definition states that the association between cue j and item i reflects the probability of the item being needed (i.e., is the target of the retrieval) given cue j:[6]

$$S_{ji} = MAS + \ln[P(i|j)] \tag{8}$$

The standard equation that calculates the fan as the number of matching items which is usually used in ACT-R implementations makes the simplifying assumption that all items associated with cue j are equally likely (i.e., *useful* in the context of cue j), such that $P(i|j) = 1/fan_j$. It is important to note here that the probability $P(i|j)$ for item i is only defined when it is associated with cue j.

In order to reflect differences in encoding strength between items (*prominence*) and cross-associations between cues, we define $P(i|j)$ here as the **match quality** Q_{ji} of item i with cue j in proportion to the match quality Q_{jv} of all active memory items v with j:

$$P(i|j) = \frac{Q_{ji}}{\sum_{v \in Items} Q_{jv}} \tag{9}$$

The next two subsections will explain how this leads to multi-associative cues and the influence of item prominence on the fan effect.

Multi-associative cues

We assume that a cue can be of variable selectivity, i.e., it can be associated with multiple features to different degrees. The general association between a cue j and a feature k is given by M_{jk}. The individual match quality Q_{ji} of cue j with a specific item i then depends on the associative strength between j and all features K_i of i.

$$Q_{ji} = \sum_{k \in K_i} M_{jk} \tag{10}$$

[6]We thank Klaus Oberauer for his helpful comments, which led to the present implementation.

As shown in Figure 5, cross-association predicts a fan effect also for items that do not share any of their features, as long as the same cue is associated with features from both items. We work through some examples next.

For the worked-out examples below, assume (see Figure 6) that an item i has feature f1 but not feature f2, and a distractor item i' has feature f2 but not f1. Assume also that the retrieval cue c1 matches f1, and cue c2 matches f2. Retrieval is triggered using the two cues c1 and c2. This is the typical target-mismatch/distractor-match scenario discussed earlier.

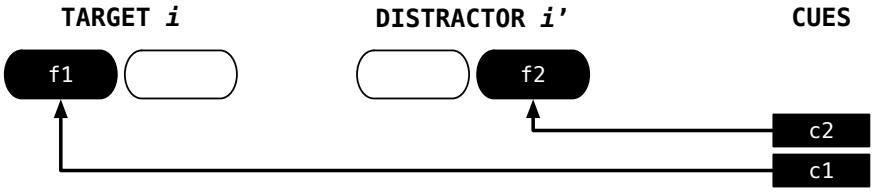


Figure 6. Standard target-mismatch/distractor-match condition without cross-associated cues.

1. **No cross-association of features (standard ACT-R case)**: In the case that there is no cross-association, the spreading activation to item i from cue c1 depends on the probability of item i given cue c1:

$$P(i|c1) = \frac{Q_{c1,i}}{\sum_{v \in Items} Q_{c1,v}} \tag{11}$$

The numerator is computed as follows. Since only feature f1 matches cue c1 in item i, we have:

$$Q_{c1,i} = \sum_{k \in K_i} M_{c1,k} = M_{c1,f1} = 1 \tag{12}$$

The denominator, $\sum_{v \in Items} Q_{c1,v}$, also has value 1 because it is the sum of the match of cue c1 to item i (which is 1) and to item i' (which is 0):

$$\sum_{v \in Items} Q_{c1,v} = Q_{c1,i} + Q_{c1,i'} = 1 + 0 \tag{13}$$

The calculation of $P(i|j)$ is therefore:

$$P(i|j) = P(i|c1) = \frac{Q_{c1,i}}{\sum_{v \in Items} Q_{c1,v}} = \frac{1}{1} = 1 \tag{14}$$

This implies that the spreading activation from cue c1 to item i is:

$$\begin{aligned} S_{c1,i} =& MAS + \ln[P(i|c1)] \\ =& MAS + \ln[1] = MAS \end{aligned} \tag{15}$$

As no other cue matches item i, $S_{c1,i}$ equals the total of spreading activation S_i that item i receives:

$$S_i = S_{c1,i} = MAS \tag{16}$$

Thus, there is no penalty to the activation of item i caused by spreading activation (fan effect) in target-mismatch/distractor-match configurations when there is no cross-association.

2. **Cross-association of 0.5**: Now consider the activation spread to item i when the cross-association level of the cues is 0.5. Under this scenario, item i receives not only 100% activation from the fully matching cue c1, but also from c2 which spreads 50% of its activation to feature f1. The distractor i' similarly gets activation not only from c2, which fully matches f2, but also from c1 which spreads 50% of its activation to feature f2. Graphically, this corresponds to the following scenario (Figure 7):

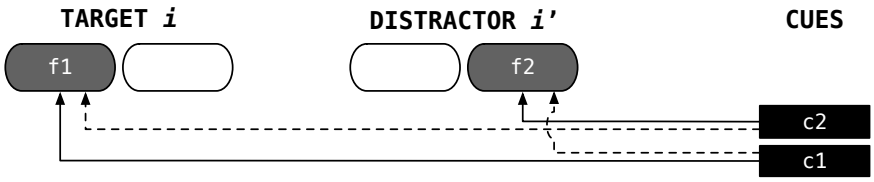


Figure 7. Target-mismatch/distractor-match condition when cues are cross-associated.

Now, $P(i|c1)$ is not 1 but 1/1.5.

$$P(i|c1) = \frac{Q_{c1,i}}{\sum\limits_{v \in Items} Q_{c1,v}} = \frac{1}{1.5} = \frac{2}{3} \tag{17}$$

This is because $Q_{c1,i} = 1$ as before, but the denominator is the sum of the match of cue c1 to item i (a match of 1) as well as the match of cue c1 to item i' (a match of 0.5).

$$\sum_{v \in Items} Q_{c1,v} = Q_{c1,i} + Q_{c1,i'} = 1 + 0.5 = 1.5 \tag{18}$$

We then use $P(i|c1)$ to calculate the spreading activation $S_{c1,i}$ from cue c1 to item i. In contrast to the scenario above without cross association, $S_{c1,i}$ now is smaller than MAS:

$$S_{c1,i} = MAS + \ln\left[\frac{2}{3}\right] = MAS + [-0.41] = MAS - 0.41 \tag{19}$$

Next, the calculation for item i and cue c2 is:

$$P(i|c2) = \frac{Q_{c2,i}}{\sum\limits_{v \in Items} Q_{c2,v}} = \frac{0.5}{1.5} = \frac{1}{3} \tag{20}$$

Here, $Q_{c2,i} = 0.5$ because of the cross-association of 0.5 of cue c2 with the feature f1. The denominator is the sum of the match of cue c2 to item i (a match of 0.5) as well as the match of cue c2 to item i' (a match of 1).

$$\sum_{v \in Items} Q_{c2,v} = Q_{c2,i} + Q_{c2,i'} = 0.5 + 1 = 1.5 \tag{21}$$

We now use $P(i|c2)$ to calculate the spreading activation $S_{c2,i}$ that item i receives from cue $c2$. Similar to $S_{c1,i}$, $S_{c2,i}$ also be smaller than MAS:

$$S_{c2,i} = MAS + \ln\left[\frac{1}{3}\right] = MAS + [-1.1] = MAS - 1.1 \tag{22}$$

Having computed $S_{c1,i}$ and $S_{c2,i}$, the term the total amount of spreading activation S_i that item i receives can be calculated (W_j is 0.5 as we have two equally weighted cues):

$$\begin{aligned} S_i =& \sum_{j \in Cues} W_j S_{ji} \\ =& \frac{1}{2} S_{c1,i} + \frac{1}{2} S_{c2,i} \\ =& \frac{1}{2}\left(MAS + \ln\left[\frac{2}{3}\right]\right) + \frac{1}{2}\left(MAS + \ln\left[\frac{1}{3}\right]\right) \\ =& MAS + \frac{1}{2}\left(\ln\left[\frac{2}{3}\right] + \ln\left[\frac{1}{3}\right]\right) \\ =& MAS + (-0.75) = MAS - 0.75 \end{aligned} \tag{23}$$

Because the spreading activation S_i received by item i will have a value less than MAS, activation of item i will go down due to the presence of the matching distractor, leading to inhibitory interference even in a target-mismatch configuration, when cross-association is greater than 0.

Prominence

We assume that the prominence of an item is reflected in its base-level activation, which also reflects how recently the item has been retrieved or created. For this purpose, we simply introduce a prominence component p_i as a constant added to the base-level activation B_i, such that Equation 1 for B_i is changed to:

$$B_i = \ln(\sum_{j=1}^{n} t_j^{-d}) + \beta_i + p_i \tag{24}$$

Thus, more prominent items are more highly activated and are therefore more likely to be retrieved. In addition, the base-level activation including prominence should affect how strongly an item interferes with the retrieval of other items: A highly activated and thus very salient item will have a stronger fan effect than an item that is less active in

memory. We therefore introduce a *saliency component* as a weighting of the individual match quality Q_{ji}, changing Equation 10 in the following way:

$$Q_{ji} = \sum_{k \in K_i} M_{jk} \times \frac{1}{1 + qe^{-(B_i - \tau)}} \tag{25}$$

The saliency component (the second factor) is a logistic function that bounds the base-level activation value between 0 and 1, such that it functions as a scaling factor for Q_{ji}. In the denominator, τ is the retrieval threshold, and q is a scaling constant that scales how strongly the match quality Q_{ji} is affected by an item's saliency. It can be used to switch the quality correction on and off and thus make our model identical to standard ACT-R: When $q = 0$, the item's base-level activation including prominence is not reflected in $P(i|j)$. Furthermore, when $q = 0$ and all cues are *maximally selective* (i.e., exactly one feature matches one cue), $P(i|j) = 1/fan_j$, in which case the model behavior is identical to standard ACT-R. If, however, $q > 0$, the base-level activation of an item—and with it the item's prominence—affects the associative strength between the retrieval cues and the item.

Figure 3 shown earlier illustrates the relationship between distractor prominence and the interference effect as predicted by the extended model, assuming that target prominence is a fix value. In addition to the facilitatory effect of highly activated distractors in target-match predicted also by standard ACT-R, the extended model additionally predicts that the fan effect only arises for sufficiently activated distractors (cf. the rising inhibition in target-match configurations in the figure).

In sum, we define the probability of a memory item i being needed given cue j, $P(i|j)$, with respect to the item's base-level activation, which in turn depends on its prominence, and its association with cue j, M_{ji}. The equations ensure that cues can be of variable selectivity (i.e., can be associated with one or more features), and that more prominent items are more strongly associated with the cues and, hence, receive more spreading activation. Since $P(i|j)$ is a probability that takes into account all memory items, both the selectivity of cues and the prominence of the item itself and of all of its competitors affect the fan effect, i.e., the strength of inhibitory interference. The equations for the total spreading activation for item i (Eq. 2) and the retrieval latency (Eq. 5) remain the same as in the original implementation.

For the simulations that we present below, we assume that factors such as syntactic position (being a subject or not) and topicalization increase the prominence and hence the base-level activation of a distractor. As the base-level activation also reflects an item's recency, the effect of interference type is predicted to add to the effect of prominence: A retroactively interfering distractor that is the discourse topic and is in subject position has the highest activation. Our account of item prominence predicts that distractor activation due to prominence and recency systematically increases the interference effect in target-match and target-mismatch configurations and can give rise to previously unexplained facilitation effects in target-match configurations for highly activated distractors.

Simulations

In order to demonstrate how item prominence and multi-associative cues change the model predictions when accounting for distractor position and different co-occurrence patterns between dependency types, we ran simulations with both the original LV05 model and the extended model with item prominence and multi-associative cues as described above. We simulated all studies that were part of the meta-analysis of Jäger et al. (2017).

Data

Figure 8 shows the number of target-match and target-mismatch comparisons that were included in our simulations, arranged by dependency type and level of distractor prominence. The data contain only studies that were part of the meta-analysis in Jäger et al. (2017). At the time, no data were available for target-match configurations in non-agreement subject-verb dependencies. A recent study by Cunnings and Sturt (2018) fills this gap but was not included in the simulations. We, however, discuss this study in the General Discussion. We categorised the experiments into three different prominence relations for the distractor: *subject position*, *topicalized*, and *other*. Subject position and topicalization are considered high prominence levels, while we do not make any a priori assumptions about which of both is more prominent than the other. The third category, *other*, stands for all relations considered low prominence, which mainly consisted of the distractor being in object position or in a prepositional phrase. As a fourth category, the figure shows the studies where the distractor was both in subject position and topicalized. We expect that the prominence in this case — and thus the distractor activation — is particularly high. As the figure shows, topicalized distractors and the combination of topicalization and subject position have so far only been tested in reflexives.

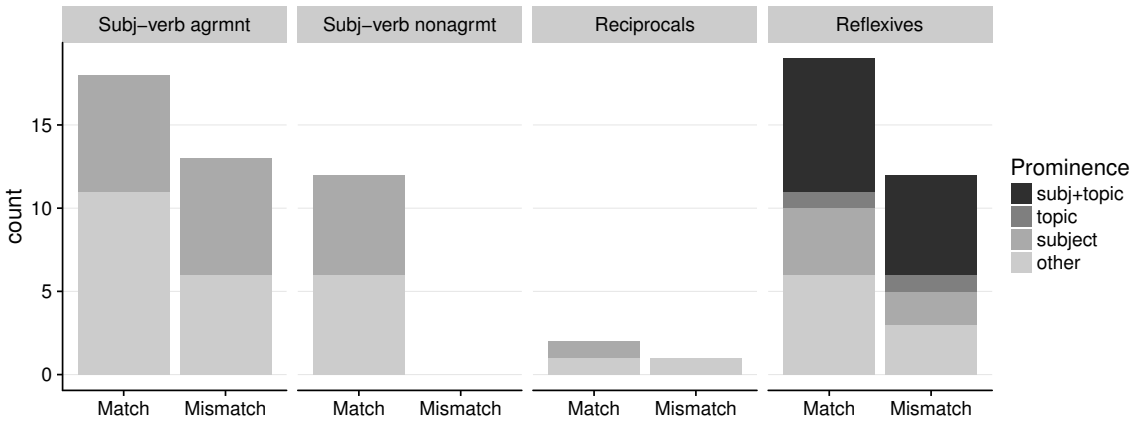


Figure 8. Number of studies included in the meta-analysis and in the simulations by dependency type and prominence.

Method

Both models were implemented in R (R Core Team, 2016). The model code is publicly available on GitHub.[7] In addition, we provide an application for running simulations with the extended model online.[8] The model simulated retrieval latencies with two or more memory items present (some studies used more than one distractor) and feature settings according to the target-match and target-mismatch conditions in Example 3. Two cues were specified at retrieval. The first cue was matched by one memory item in all conditions, which distinguished this item as the target. The second cue was matched by the target in conditions a and b (target-match) and by the distractor in conditions b and c (distractor-match). In order to ensure common parameter settings within experiments, the 77 data points used in the meta-analysis were modeled in 51 *experimental sets*, such that parameters were held constant between target-match and target-mismatch conditions of the same experiment.

Parameter estimation. As is common practice in ACT-R modeling, we estimated the latency factor F (see Equation 5) for each experiment in both models to scale the predictions into a range that is comparable with the data. For the simulations with the extended model, we estimated different values of the distractor prominence parameter p_{dstr} for each of three prominence categories within dependency types: *low* (neither subject nor topic), *medium* (subject *or* topic), and *high* (subject *and* topic). The parameter p_{dstr} was restricted according to the experimental design to three ordered ranges of values such that medium prominence was constrained to be close to the target prominence ($p_{trgt} = 0$) with $-1 \leq p_{dstr} \leq 2$, whereas low prominence was constrained to be smaller than the target prominence ($-3 \leq p_{dstr} \leq 0$), and high prominence was bound to higher values ($1 \leq p_{dstr} \leq 4$). Thus, the full range of predictions shown in Figure 3 can be generated theoretically, but the generating process is restricted to specific properties of the distractor. We allowed the value ranges to overlap, because we had no specific assumptions about the prominence parameter values except for the ordering of the levels. Since the relevant manipulation was the difference between distractor prominence and target prominence, only distractor prominence was estimated and target prominence stayed fixed at the default prominence value of 0.

The cross-association level c was estimated only for the two cases motivated above: reciprocals and the Chinese reflexive *ziji*. It was set to 0 otherwise.

Interference type (retro vs. proactive interference) was reflected in the model by manipulating the order of target and distractor. For retroactive designs, the target was more distant from the retrieval site than the distractor and vice versa. Hence, interference type affects the model through the memory decay component, which reduces the activation of an item as a function of time.

Results

We ran simulations with the LV05 model and the extended model. Because Lewis and Vasishth (2005) speculated that model fit might improve without the decay component

[7] The model code is available at https://github.com/felixengelmann/inter-act.

[8] The application was built using the *shiny* package (Chang, Cheng, Allaire, Xie, & McPherson, 2016) and is located at https://engelmann.shinyapps.io/inter-act/.

of ACT-R,[9] we also ran variants of both models without the decay component. Interference effects were computed within target-match and target-mismatch conditions as the difference between distractor-match (high interference) and distractor-mismatch (low interference) conditions, averaged over 5000 iterations per simulation.

Table 3
Estimated values for prominence parameter p_{dstr} in extended model with decay for three prominence levels.

Dependency	low	medium	high
agreement	0.00	1.50	
nonagreement	0.00	0.00	
reci	-2.00	0.50	
refl	-1.50	-1.00	4.00

Table 4
Root-mean-square deviation between model predictions and observed data, averaged within dependency type and model (best values in bold).

Dependency	LV05	LV05 (no decay)	Extended	Extended (no decay)
Subj-verb agreement	18.01	15.65	14.67	**13.29**
Subj-verb non-agrmnt	6.97	8.14	**5.18**	7.81
Reflexives/Reciproc.	12.44	11.79	7.5	**6.35**

Distractor prominence values p_{dstr} were estimated for the three prominence categories low (neither subject or topic), medium (subject or topic), and high (subject and topic). The resulting estimates are shown in Table 3. Table 4 summarizes the fit of all four model configurations, averaged within dependency type. Overall, the extended model fit the available data better than the original model of LV05. Except for non-agreement subject-verb dependencies, the use of decay did not improve the fit with the data. With respect to the extended model, decay only improved the fit for non-agreement subject-verb dependencies but, for the other dependency types, produced a worse fit compared to the model without decay. Since decay generally does not improve the fit, this suggests that the information about the linear order of target and distractor (pro- vs. retroactive interference) may not be useful as a predictor in the models considered here. We revisit this point in the General Discussion.

Figure 9 illustrates the average patterns of effects across prominence categories within dependency types, comparing the effect means between the data and the two models.

In *subject-verb agreement dependencies*, the prominence parameter p_{dstr} in the extended model (Table 3) was estimated to be higher for medium prominence experiments compared to low prominence experiments. This leads to the prediction that, when prominence increases, the target-match inhibitory interference effect decreases, and the target-

[9] Lewis and Vasishth (2005) write on p. 408: "Any structural or quantitative change to the model that moves in the direction of decreased emphasis on decay and increased emphasis on interference would likely yield better fits."

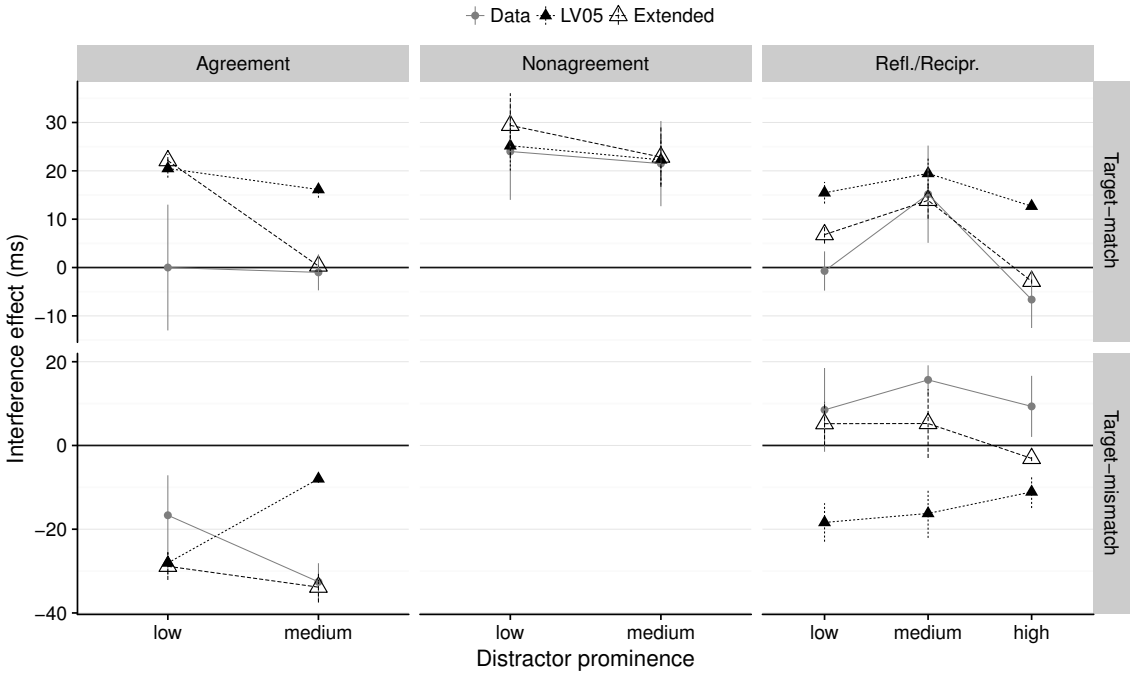


Figure 9. Mean target-match and target-mismatch effects in the data and predictions of the LV05 and extended models by distractor prominence level within dependency types (subject-verb agreement, non-agreement, reciprocals/reflexives).

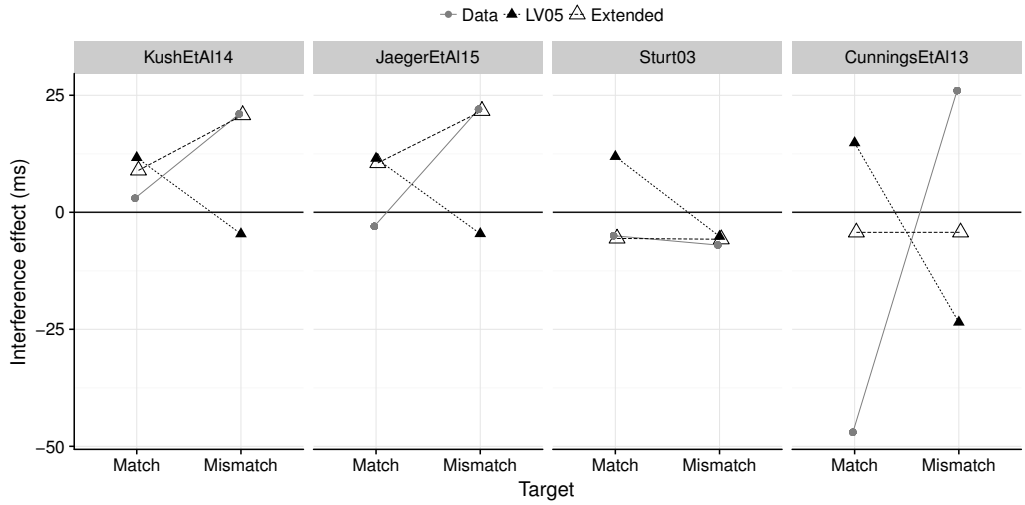


Figure 10. Data and predictions of LV05 and the extended model for interference effects of Kush and Phillips (2014), Jäger et al. (2015, Exp. 1), Sturt (2003, Exp. 1), and Cunnings and Felser (2013, Exp. 2, low working memory).

mismatch facilitatory effect increases. This prediction fits the pattern in the data better than the original model LV05.

For *non-agreement subject-verb dependencies*, the fit did not improve in the extended model, because the data only contain target-match configurations, for which the results — mainly inhibitory interference — are perfectly compatible with LV05. There are also no differences between prominence categories in the data. Consequently, the prominence parameter was not estimated to be different between low and medium prominent distractors (Table 3).

The biggest improvements in comparison with the original model were achieved for *reflexive and reciprocal dependencies*. As can be seen in Figure 9, the average effects in target-match configurations show increasing inhibition from low to medium prominence and facilitatory interference in high prominence. This is exactly the pattern that our prominence model predicts (see Figure 3 shown earlier). Consequently, the extended model fits the data better than LV05, and also predicts a facilitatory effect for highly prominent distractors on average. In target-mismatch configurations, the data shows inhibitory effects on average in all three prominence categories. This is incompatible with LV05. However, the extended model approximately predicts this pattern as a consequence of just two studies (Kush & Phillips, 2014 and Jäger et al., 2015) that are simulated with cross-associated cues due to their linguistic contexts of reciprocals and Chinese reflexives.

Figure 10 shows the data and simulation results for four exemplary cases where the data *qualitatively* deviates from the predictions of the original LV05 model. The studies by Kush and Phillips (2014) on reciprocals and by Jäger et al. (2015) on Chinese reflexives are two cases of low feature-selectivity as explained in the section on multi-associative cues. As a result of the cue-feature cross-association, the extended model predicts *inhibitory* interference effects in target-mismatch configurations, whereas LV05 predicts facilitation. The model parameter for the cross-association level was estimated at 1 for reciprocals (Kush & Phillips, 2014) and at 0.9 for *ziji* (Jäger et al., 2015).

Cunnings and Felser (2013), and Sturt (2003) are examples of *facilitatory* effects in target-match configurations, which only the extended model accounts for as a consequence of high distractor prominence values.

However, Cunnings and Felser (2013) is also an example of a pattern that is not compatible with any of the two tested models. The inhibitory target-mismatch effect is not predicted by the extended model, because no increased cross-association is assumed in English reflexives. And even if the cross-association level was assumed to be elevated in this case, it would be impossible to predict an inhibitory target-mismatch effect and a facilitatory target-match effect at the same time. Hence, under the assumptions of the two cue-based retrieval models tested here, the data of Cunnings and Felser (2013) are unexpected. We return to this point in the General Discussion.

General Discussion

The aim of this work was to show the quantitative constraints of the Lewis and Vasishth (2005) model and investigate the consequences of memory accessibility and context-dependent cue-feature associations in the light of the available evidence on interference effects in dependency resolution. We have presented an implemented account of item prominence and multi-associative cues as an extension to the cue-based retrieval model of LV05.

Our simulations show that item prominence and multi-associative cues predict a range of data points that were previously not predicted by the model. This suggests that the assumptions of the original LV05 model were not entirely correct: it is important to account for different aspects of memory accessibility, for individual study design, and context-based feature-selectivity in order to generate accurate predictions of a model of cue-based memory retrieval such as LV05. We therefore believe that these independently motivated extensions help to more precisely interpret individual empirical results as being evidence in favor of or against the model. The simulations presented here thus provide new insights into the cognitive mechanisms behind interference effects.

The model comparisons also suggest that decay could play a smaller role than generally assumed. Indeed, independent work in psychology argues that interference rather than decay is the more important construct (Berman, Jonides, & Lewis, 2009; Lewis & Badecker, 2010; Oberauer & Lewandowsky, 2013, 2014). However, we cannot conclusively say whether decay has no impact or is only disguised by a counteracting effect of prominence. This is because interference type (pro- vs. retroactive interference) and distractor prominence are confounded in the literature: Studies with prominent distractors more often used a proactive rather than a retroactive interference design, whereas studies with non-prominent distractors more often used a retroactive interference design (see Table 5 in the Appendix). Hence, the two factors prominence and interference type, which both influence the distractor activation in memory, might tend to cancel each other out due to experimental design. The role of decay could be investigated in future work by designing an experiment that crosses pro- and retroactive distractor position with the prominence of the distractor.

The predictions of the model are severely restricted by the fact that the same cognitive mechanisms and the same parameter values are assumed for simulating both target-match and target-mismatch configurations within a given experiment. This restricts the predictions of the model considerably; for example, the model cannot predict, for a given experiment, an *inhibitory* effect in target-mismatch as well as a *facilitatory* effect in target-match configurations, which was found in gaze durations of readers with low working memory capacity by Exp. 2 of Cunnings and Felser (2013) as shown in Figure 10. This is because a facilitatory target-match effect is caused by a high distractor activation that overrides the fan effect. Consequently, the fan effect must be eliminated in both the target-match *and* target-mismatch configurations in the presence of a highly prominent distractor even if we assumed a high cross-association level. Hence, the model makes the strong prediction that the pattern observed by Cunnings and Felser (2013) should not occur. If the model simulations had involved separate parameter fits for target-match and mismatch within the same experiment, the model would have been able to predict this and other patterns that are implausible under the model's cognitive assumptions. Our simulations therefore considerably restrict the model's prediction space. If any outcome were possible, the model would not be very useful.

It is important to reiterate that the extended model is not an ad-hoc model aiming at an improved fit with the available data but that the extensions are based on independently motivated assumptions. The comparison of the model predictions with the data thus not only tests the model but at the same time helps to interpret the available data. As the meta-analysis (Jäger et al., 2017) points out, low power and publication bias could be important

factors that weaken the empirical base. For example, Appendix B of Jäger et al. (2017) shows that power for many of the published studies on interference could be as low as 10-20%. As Gelman and Carlin (2014) have pointed out, low-power studies will not only fail to detect an effect under repeated sampling, but when an effect is found to be significant, it can often have the wrong sign (Type S error) and/or be greatly exaggerated in magnitude (Type M error). It would therefore be worthwhile to re-evaluate the predictions of this extended LV05 model with larger-sample studies.

The data on non-agreement subject-verb dependencies concurs overall with the general LV05 predictions — inhibition in target-match configurations — and was thus fit well by both models. The picture is, however, incomplete since no data on target-mismatch configurations for this dependency type were available at the time of the Jäger et al. (2017) meta-analysis and thus not included in our simulations. However, a recent study by (Cunnings & Sturt, 2018) showed evidence for a facilitatory effect in target-mismatch configurations in non-agreement subject-verb dependencies, which is predicted by LV05. They conducted two eyetracking while reading studies in which they manipulated the plausibility of the correct dependent of the verb, and the plausibility of the distractor noun. They showed that when the correct dependent is implausible, the distractor's plausibility influences reading time at the verb, such that a facilitation is observed. For example, faster total reading times were observed at the verb *shattered* in 8a compared to 8b. Our own Bayesian estimate of their effect size is -22 ms with a credible interval of $[-4, -42]$.

(8) a. Sue remembered the letter that the butler with the cup accidentally shattered today in the dining room.

b. Sue remembered the letter that the butler with the tie accidentally shattered today in the dining room.

A major contribution of the present work is that it spells out, for the first time, the predictions of the LV05 model with reference all the evidence available. The modeling presented here is highly constrained: (i) The presented model is built on independently motivated — and, in terms of ACT-R, domain-independently validated — assumptions about memory retrieval, item prominence, and multi-associative cues, which are sensitive to experimental design choices; (ii) the model predictions are restricted by interactions between variables such as prominence, recency, and cue-feature cross-association; and (iii) the parameters are fixed within a given experiment, thus ruling out certain patterns of target-match and target-mismatch effects. An important prediction of the model in this respect is that the previously unexplained observations of facilitation in target-match or inhibition in target-mismatch can be explained under certain conditions, but, as explained above, seeing both *in the same experiment* is impossible according to the model assumptions. Constrained predictions such as these are important because they make the theory falsifiable in principle.

As we have discussed above, the conclusions to be drawn about prominence and cue associations are preliminary because (i) the available data are sparse with respect to the levels of distractor prominence studied within dependency types and different levels of feature selectivity, (ii) there may be confounds between prominence and other factors, and (iii) there may be different cognitive processes involved in certain dependency types that

the model does not account for. In the following, we further discuss the implications of our approach for distractor prominence and cue-feature associations and potential alternatives.

Distractor activation

In the model we have presented, the prominence of a distractor is a function of its syntactic position and discourse saliency. An alternative account of how distractor position could affect the magnitude of interference has been discussed in Van Dyke and McElree (2011). By way of a weighting mechanism, a mismatching syntactic feature would lower the consideration of a distractor as a retrieval candidate—or, with gating rather than weighting, even rule it out completely, irrespective of any matching semantic or pragmatic features. This account predicts that interference effects are very small or absent if a distractor does not match the syntactic requirement, e.g., of being a grammatical subject. The predictions of syntactic weighting are consistent with our prominence account and are also compatible with ACT-R and LV05. Because of its reduced activation, a distractor that mismatches the subject would have a very low probability of being retrieved instead of the target, and, thus, no facilitatory interference is expected in target-mismatch configurations. The fan effect in target-match configurations would not be directly affected, because the fan effect in ACT-R is a consequence only of the feature that is manipulated between two conditions: The difference in the target activation between the distractor-match and the distractor-mismatch conditions is the same no matter how many additional cues the distractor matches across conditions. However, an effect of syntactic match in target-match configurations would nevertheless be predicted on the basis of a generally lower activated target: Because the relation between activation and latency in ACT-R is a negative exponential function (cf. Equation 5), differences in activation have less impact on the retrieval speed for items with a higher activation than for items with a lower activation. In case distractor and target both match the subject cue, the fan effect reduces both in activation across conditions compared to the case when only the target matches the subject cue. As a consequence, when the distractor matches the subject cue, the retrieval latency of the target is more affected by the fan effect of a feature manipulation, i.e., a greater inhibitory interference effect is predicted in target-match configurations.

Hence, the predictions of the syntactic weighting account regarding syntactic position are similar to the predictions of our prominence account: A distractor in subject position compared to object position increases the inhibitory interference effect in target-match configurations and the facilitatory effect in target-mismatch configurations. However, the predictions of syntactic weighting are only valid when it can be assumed that grammatical position is part of the retrieval cues. In contrast, the predictions of our prominence account are independent of cue combinatorics and the match quality of the distractor at retrieval. Instead, the predictions rest on the assumption that items in subject position have a higher relevance for interpreting a sentence and are, thus, maintained more actively in memory (Brennan, 1995; Chafe, 1976; Grosz et al., 1995; Keenan & Comrie, 1977). In the same way, this account of prominence due to relevance can be extended to discourse saliency such as topic or other contributing factors that we have not considered here: For example, thematic role (J. E. Arnold, 2001), contrastive focus (Cowles, Walenski, & Kluender, 2007), first mention (Gernsbacher & Hargreaves, 1988), and animacy (Fukumura & van Gompel, 2011) are known to affect discourse saliency and might thus influence distractor promi-

nence. Furthermore, a facilitatory effect in target-match configurations as a consequence of distractor prominence cannot be explained in terms of cue combinatorics.

Multi-associative cues

The principle of multi-associative cues states that cues can be associated with multiple features to different degrees depending on experience with the linguistic context. Crossed cue-feature associations between two cues predict inhibitory interference in target-mismatch conditions for dependency environments with high feature-co-occurrence in comparison to environments with low feature-co-occurrence. This is based on the assumption that cue-feature associations are the result of associative learning through exposure to different dependency types and their grammatical antecedents. The learning process would be best described along the lines of the *naive discriminative learning* model developed by Baayen, Milin, Đurđević, Hendrix, and Marelli (2011). Their model is an implementation of the Rescorla & Wagner equations for classical conditioning based on the presence and absence of cues and outcomes and has been applied to a range of effects in the context of language acquisition.

A possible way to test the multi-associative cues hypothesis for English in a controlled experiment would be to directly compare reflexives and reciprocals, manipulating the number cue in both. An example design we have also suggested in Jäger et al. (2015) is shown in Example 9.

(9) a. *Reflexive; distractor-match*
The *nurse* who cared for the *children* had pricked *themselves* ...

b. *Reflexive; distractor-mismatch*
The *nurse* who cared for the *child* had pricked *themselves* ...

c. *Reciprocal; distractor-match*
The *nurse* who cared for the *children* had pricked *each other* ...

d. *Reciprocal; distractor-mismatch*
The *nurse* who cared for the *child* had pricked *each other* ...

Under the multi-associative cues hypothesis, a reduced facilitatory effect or an inhibitory effect is predicted for the reciprocal *each other* compared to the reflexive *themselves*. In order to derive a finer-grained metric that predicts differences in cue-feature cross-association levels between different dependency environments, co-occurrence frequencies could be computed from a corpus in which sufficient dependency information is available.

Our theory of multi-associative cues predicts a higher cross-association level for both reciprocals and the Chinese reflexive *ziji* compared to English reflexives. This could explain the result of Kush and Phillips (2014), who found inhibitory interference in target-mismatch conditions in Hindi reciprocals, as well as our finding of an inhibitory target-mismatch effect for *ziji* in Experiment 1 of Jäger et al. (2015). The modeling results (Figure 9) showed that these two studies were sufficient to cause the average target-mismatch effect to be inhibitory in low and medium prominence reflexive/reciprocal studies. According to the meta-analysis in Jäger et al. (2017), the overall interference effect in target-mismatch configurations studies of reflexive- and reciprocal-antecedent dependencies is inhibitory (see

Table 1). Importantly, this overall inhibitory effect was found even when excluding the Chinese reflexives study of Jäger et al. (2015), which had a larger-than-usual sample size and could therefore have unduly influenced the meta-analysis. Due to the two studies with cross-associated cues, the extended model predicted a tendency for an inhibitory effect on average in target-mismatch configurations, but not as clear as the meta-analysis found. A less conservative simulation with a freely varying cross-association parameter would, however, result in an overall increased cross-association level for reflexives compared to subject-verb agreement dependencies (subject-verb agreement showed an overall facilitatory effect in target-mismatch configurations). In support for a theory of higher feature-co-occurrence and, thus, a higher cross-association level in reflexive-antecedent than in subject-verb dependencies in general, one could argue that reflexive-antecedent dependencies have a rather restrictive set of cues that define the target, whereas subject-verb dependencies occur in a wide range of contexts in which various semantic cues in addition to grammatical ones might be used (cf. Van Dyke & McElree, 2006).

Under a theory of multi-associative cues, an interesting question is whether categorically distinguishing two cues requires cognitive effort. If so, one would expect an additional variation of the cross-association level that depends on task demands and individual differences. There is evidence that the depth of linguistic processing is influenced by task-specification (Logačev & Vasishth, 2016; Swets, Desmet, Clifton, & Ferreira, 2008) and individual differences (von der Malsburg & Vasishth, 2013; Nicenboim et al., 2016; Traxler, 2007), resulting in underspecification of sentence representations or "good-enough processing" (Ferreira, Ferraro, & Bailey, 2002). In the same way, multiple cue-feature associations could be part of a dynamically adapted resource-preserving strategy. This assumption predicts elevated cross-association levels for readers with less cognitive resources in order to compensate for slower processing. It also predicts increased cross-association for experiments with little task demand, like easy comprehension questions, because the effort of a precise cue specification would not be necessary. There is one experiment on reflexives that controlled for participants' working memory capacity: Cunnings and Felser (2013) found in their Experiment 2 on English reflexives an inhibitory effect on the critical region in target-mismatch conditions only for low-capacity readers. The effect is only marginally significant (mean 22 ms, SE 26 ms) but would be in line with the assumption of an individual-level variation of cue-feature associations due to adaptive processes. Note, however, that, even if it was the case that low-capacity readers experience higher cross-association, for reasons explained above, the current model could not predict an *inhibitory* target-mismatch effect at the same time as a *facilitatory* target-match effect as is the case in Cunnings & Felser, 2013. Since there is only one experiment testing low-capacity readers on target-mismatch configurations, a hypothesis of cue-feature associations being adaptive to individual capacity limits is currently speculative, and planned experiments are needed in order to test it.

Other factors besides feature-co-occurrence that affect the strength of cue associations have not been considered here. Most prominently, it has been claimed that syntactic cues are weighted more strongly than semantic cues (e.g., Nicol, 1988; Sturt, 2003; Van Dyke, 2007; Van Dyke & McElree, 2011). A stronger weighting for syntactic cues might actually be subsumed by co-occurrence, assuming that syntactic cues are more reliable (i.e., have a higher co-occurrence) in a certain construction than semantic cues.

Other associations may, however, go beyond pure co-occurrence. For example, an experiment conducted by Van Dyke and McElree (2006) showed interference effects based on similarities between nouns that tap into world knowledge, such as the property of being *fixable*. Some cues may be stronger than others based on their semantics and pragmatics: Carminati (2005) have proposed a hierarchy between features, such that *person* > *number* > *gender*. Additionally, in English, *number* is morphologically overt while *animacy* and *gender* are not. The effects of semantically, pragmatically, or morphologically motivated differences between retrieval cues remain to be investigated.

Conclusion

The extended model of cue-based retrieval provides, for the first time, quantitative predictions with respect to systematic variability in experimental design across studies. The presented model is therefore an important step forward in helping us interpret results in the context of previous findings and for formulating computationally informed predictions for future experiments.

The two principles of item prominence and multi-associative cues that constitute our extended model are compatible with the general ACT-R theory of cue-based retrieval as the essential mechanism underlying dependency resolution in sentence processing. Both principles are independently motivated and should be considered as domain-general mechanisms and as extensions to the current ACT-R architecture. Looking beyond ACT-R, future work should also investigate whether inhibitory interference in target-mismatch configurations can be explained in terms of other computational/mathematical models of memory, such as the well-known drift-diffusion model account of Ratcliff (1978). A further, very productive line of inquiry would be a systematic study of the quantitative predictions of other computational models of dependency completion in language comprehension (Cho et al., 2017; Rasmussen & Schuler, 2017; Smith et al., 2018) relative to published data.

Researchers are invited to use the extended model presented here to conduct further simulations. In order to facilitate this, we provide an online application of the model at https://engelmann.shinyapps.io/inter-act.

Acknowledgements

For valuable comments and discussions, we are grateful to Colin Phillips, Hedderik van Rijn, Niels Taatgen, Patrick Sturt, Ian Cunnings, Brian Dillon, Dan Parker, Dave Kush, Sol Lago, Bruno Nicenboim, and Dario Paape. We also want to thank the audiences of AMLaP 2014, CUNY 2015, and ICCM 2015 for commenting on our work. This research was partly funded by a Volkswagen Foundation grant (Grant Number 89 953) to Shravan Vasishth, and by the SFB 1287, Limits of Variability in Language, project Q, which funded Lena Jäger (PIs: Shravan Vasishth and Ralf Engbert) and B2 (PIs: Ralf Engbert and Shravan Vasishth).

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive psychology*, *6*(4), 451–474.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–60.

Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ariel, M. (1990). *Accessing noun-phrase antecedents.* London, UK: Routledge.

Arnold, D. (2007). Non-restrictive relatives are not orphans. *Journal of Linguistics*, *43*(2), 271–309.

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*, 137–162.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748–769.

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 317.

Boston, M., Hale, J., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*(3), 301–349.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*(2), 137–167.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733.

Carminati, M. N. (2005). Processing reflexes of the feature hierarchy (person> number> gender) and implications for linguistic theory. *Lingua*, *115*(3), 259–285.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–56). New York: Academic Press.

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2016). shiny: Web Application Framework for R [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=shiny (R package version 0.14)

Chen, Z., Jäger, L. A., & Vasishth, S. (2012). How structure-sensitive is the parser? Evidence from Mandarin Chinese. In B. Stolterfoht & S. Featherston (Eds.), *Empirical approaches to linguistic theory: Studies of meaning and structure* (pp. 43–62). Berlin, Germany: Mouton de Gruyter.

Cho, P. W., Goldrick, M., & Smolensky, P. (2017). Incremental parsing in a continuous dynamical system: Sentence processing in Gradient Symbolic Computation. *Linguistics Vanguard*, *3*(1).

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

Cowles, H. W., Walenski, M., & Kluender, R. (2007). Linguistic and cognitive prominence in anaphor resolution: Topic, contrastive focus and pronouns. *Topoi*, *26*, 3–18.

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*(1-2), 188–219.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, *75*, 117–139.

Cunnings, I., & Sturt, P. (2018). *Retrieval interference and sentence interpretation.* (Under review)

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.

Du Bois, J. W. (2003). Argument structure: Grammar in use. In J. W. Du Bois, L. E. Kumpf, & W. J. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function* (Vol. 14, pp. 11–60). Amsterdam, The Netherlands: John Benjamins.

Engelmann, F. (2016). *Toward an integrated model of sentence processing in reading* (Unpublished doctoral dissertation). University of Potsdam, Potsdam, Germany.

Engelmann, F., Vasishth, S., Engbert, R., & Kliegl, R. (2013). A framework for modeling the interaction of syntactic processing and eye movement control. *Topics in Cognitive Science*, *5*(3), 452-474.

Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, *12*, 485–502.

Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.

Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in Psychology*, *6*(349).

Fukumura, K., & van Gompel, R. P. G. (2011). The effect of animacy on the choice of referring expression. *Language and Cognitive Processes*, *26*(10), 1472–1504.

Gelman, A., & Carlin, J. (2014). Beyond power calculations assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*, 699–717.

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203–225.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, *6*(617).

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

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, *8*(1), 63–99.

King, J., Andrews, C., & Wagers, M. (2012). Do reflexives always find a grammatical antecedent for themselves? In *25th Annual CUNY Conference on Human Sentence Processing* (p. 67). New York, NY: The CUNY Graduate Center.

Kush, D. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology*, *5*(1252).

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.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford University Press.

Lewis, R. L., & Badecker, W. (2010). Short-term memory in sentence production and its adaptive control. In *The CUNY conference of human sentence processing*.

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. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

Logačev, P., & Vasishth, S. (2016). A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cognitive Science*, *40*(2), 266–298.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578.

Mätzig, P., Vasishth, S., Engelmann, F., Caplan, D., & Burchert, F. (2018). A computational investigation of sources of variability in sentence comprehension difficulty in aphasia. *Topics in Cognitive Science*. (accepted)

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course

analysis. *Journal of Memory and Language*, *32*, 536–571.
McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*(2), 111–123.
McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*, 67–91.
Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory differences in unbounded dependencies. *Frontiers in Psychology*, *7*(280). (Special Issue on Encoding and Navigating Linguistic Representations in Memory) doi: 10.3389/fpsyg.2016.00280
Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language*. doi: 10.1016/j.jml.2017.08.004
Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in german. *Cognitive Science*. (accepted) doi: 10.1111/cogs.12589
Nicol, J. (1988). *Coreference processing during sentence comprehension* (PhD thesis). Massachusetts Institute of Technology, Cambridge, MA.
Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. *Journal of Experimental Psychology: General*, *142*(2), 380.
Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language*, *73*, 15 - 30.
Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, *157*, 321–339.
Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language*, *94*, 272–290.
Patil, U., Hanne, S., Burchert, F., De Bleser, R., & Vasishth, S. (2016). A computational evaluation of sentence comprehension deficits in aphasia. *Cognitive Science*, *40*, 5–50. doi: 10.1111/cogs.12250
Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology*, *7*(329). (Special Issue on Encoding and Navigating Linguistic Representations in Memory)
Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456.
R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/
Rasmussen, N. E., & Schuler, W. (2017). Left-corner parsing with distributed associative memory produces surprisal and locality effects. *Cognitive science*.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59.
Reinhart, T. M. (1976). *The syntactic domain of anaphora* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (Vol. 2, pp. 64–99). New York, NY: Appleton-Century-Crofts.
Roberts, S., & Pashler, H. (2000, April). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358-367.
Schneider, D. W., & Anderson, J. R. (2012). Modeling fan effects on the time course of associative recognition. *Cognitive Psychology*, *64*(3), 127–160.
Smith, G., Franck, J., & Tabor, W. (2018). *Semantic features unpack notional plurality in pseudopartitive agreement: A self-organizing approach.* (Manuscript under revision, Cognitive Science)

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*, 542–562.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory and Cognition*, *36*(1), 201–216.

Tomasello, M. (2003). Constructing a language: A usage-based account of language acquisition. *Cambridge, MA*.

Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory and Cognition*, *35*(5), 1107–1121.

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*(347).

Van Dyke, J. (2002). *Parsing as working memory retrieval: Interference, decay, and priming effects in long distance attachment* (Unpublished doctoral dissertation). University of Pittsburgh, PA.

Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(2), 407–430.

Van Dyke, J., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*, 285–316.

Van Dyke, J., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.

Van Dyke, J., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*(2), 225–258.

Vasishth, S., Bruessow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4).

Vasishth, S., & Engelmann, F. (to appear). *Sentence comprehension as a cognitive process: A computational approach.* Retrieved from https://vasishth.github.io/sccp/

Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *104*(4), 442–452.

Table 5
List of experiments included in the simulations.

Dependency	Prominence	ID	Publication	Int. type	Lang.	Distr. pos.
S-V agreement	low	1	Franck et al. (2015, E1, Compl)	pro	FR	obj
		2	Franck et al. (2015, E1, RC)	pro	FR	obj
		3	Dillon et al. (2013, E1)	retro	EN	obj
		4	Pearlmutter et al. (1999, E1)	retro	EN	PP
		5	Pearlmutter et al. (1999, E2)	retro	EN	PP
		6	Pearlmutter et al. (1999, E3, plur)	retro	EN	PP
		7	Pearlmutter et al. (1999, E3, sing)	retro	EN	PP
		8	Tucker et al. (2015)	retro	AR	obj
		9	Wagers et al. (2009, E4, PP)	retro	EN	PP
		10	Wagers et al. (2009, E5)	retro	EN	PP
		11	Wagers et al. (2009, E6)	retro	EN	PP
	medium	12	Lago et al. (2015, E1)	pro	SP	subj
		13	Lago et al. (2015, E2)	pro	EN	subj
		14	Lago et al. (2015, E3a)	pro	SP	subj
		15	Lago et al. (2015, E3b)	pro	SP	subj
		16	Wagers et al. (2009, E2)	pro	EN	subj
		17	Wagers et al. (2009, E3, RN, plur)	pro	EN	subj
		18	Wagers et al. (2009, E3, RN, sing)	pro	EN	subj
S-V non-agrmnt	low	19	VanDyke et al. (2006)	pro	EN	3x memory
		20	VanDyke et al. (2011,E2b)	pro	EN	obj
		21	VanDyke (2007, E1, LoSyn)	retro	EN	PP
		22	VanDyke (2007, E3, LoSyn)	retro	EN	PP
		23	VanDyke (2007, E2, LoSyn)	retro	EN	PP
		24	VanDyke et al. (2011, E2b)	retro	EN	obj
	medium	25	VanDyke et al. (11E1bpro)	pro	EN	subj
		26	VanDyke et al. (11E1bretro)	pro	EN	subj
		27	VanDyke (2007, E1, LoSem)	retro	EN	PP, subj
		28	VanDyke (2007, E2, LoSem)	retro	EN	PP, subj
		29	VanDyke (2007, E3, LoSem)	retro	EN	PP, subj
		30	VanDyke et al. (2003, E4)	retro	EN	PP, subj
Reciprocals	low	31	Kush et al. (2014)	retro	HI	prepobj
	medium	32	Badecker et al. (2002, E4)	pro	EN	subj
Reflexives	low	33	Badecker et al. (2002, E5)	pro	EN	gen
		34	Badecker et al. (2002, E6)	pro	EN	prepobj
		35	Jäger et al. (2015, E2)	pro	CN	3x memory
		36	Dillon et al. (2013, E1)	retro	EN	obj
		37	Dillon et al. (2013, E2a)	retro	EN	obj
		38	Dillon et al. (2013, E2b)	retro	EN	obj
	medium	39	Badecker et al. (2002, E3)	pro	EN	subj
		40	Chen et al. (2012, local)	retro	CN	subj
		41	Jäger et al. (2015, E1)	retro	CN	subj
		42	Patil et al. (2016)	retro	EN	subj
		43	Sturt (2003, E2)	retro	EN	obj, topic
	high	44	Cunnings et al. (2013, E1, HiWMC)	pro	EN	subj, topic
		45	Cunnings et al. (2013, E1, LoWMC)	pro	EN	subj, topic
		46	Cunnings et al. (2014, E1)	pro	EN	subj, topic
		47	Felser et al. (2009, inaccMism)	pro	EN	subj, topic
		48	Felser et al. (2009, noCcom)	pro	EN	subj, topic
		49	Sturt (2003, E1)	pro	EN	subj, topic
		50	Cunnings et al. (2013, E2, HiWMC)	retro	EN	subj, topic
		51	Cunnings et al. (2013, E2, LoWMC)	retro	EN	subj, topic

Note. The experiments are ordered by dependency type, prominence level, and interference type. The experiments are further classified by language (AR = Arabic, CN = Mandarin Chinese, EN = English, FR = French, HI = Hindi, SP = Spanish) and by syntactic position of the distractor (subject, object, genitive attribute, prepositional phrase, sentence external memory load, discourse topic).