COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 43 (2019) e12800

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12800

# The Effect of Prominence and Cue Association on Retrieval Processes: A Computational Account

Felix Engelmann,[a] Lena A. Jäger,[b] Shravan Vasishth[c]

[a]*Development and Hearing, Division of Human Communication, University of Manchester*
[b]*Department of Computer Science, University of Potsdam*
[c]*Department of Linguistics, University of Potsdam*

Received 4 January 2018; received in revised form 3 October 2019; accepted 10 October 2019

**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). The comparison shows that the model has only partial success in explaining the data; and we propose that its prediction space is restricted by oversimplifying assumptions. We then implement a revised model that takes into account differences between individual experimental designs in terms of the prominence of the target and the distractor in memory- and context-dependent cue-feature associations. The predictions of the original and the revised model are quantitatively compared with the results of the meta-analysis. Our simulations show that, compared to the original LV05 model, the revised model accounts for the data better. The results suggest that effects of prominence and variable cue-feature associations need to be considered in the interpretation of existing empirical results and in the design and planning of future experiments. With regard to retrieval interference in sentence processing and to the broader field of psycholinguistic studies, we conclude that well-specified models in tandem with high-powered experiments are needed in order to uncover the underlying cognitive processes.

*Keywords:* ACT-R; Cue-based retrieval; Sentence processing; Dependency completion; Retrieval interference; Computational modeling

Correspondence should be sent to Felix Engelmann, LuCiD Child Study Centre, Coupland 1, University of Manchester, Oxford Road, Manchester M13 9PL, UK. E-mail: felix.engelmann@posteo.net

## 1. 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 *features* 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 process 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 each 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, lowering retrieval accuracy and increasing retrieval time at the verb phrase *was complaining*. This kind of similarity-based interference is called inhibitory interference.

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 antecedent noun phrase to resolve its reference; this antecedent, the stereotypically masculine noun phrase *the tough soldier*, must be the subject of the main clause because English has a constraint (Principle A of the binding theory, Chomsky, 1981) that requires the antecedent to be inside the reflexive's binding domain (i.e., roughly speaking, it needs to be in the same clause) as the reflexive and in a particular syntactic relation (called c-command, Reinhart, 1976) with respect to the reflexive. 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 subject, cannot be the antecedent of the reflexive *himself*: 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, masculine gender marking. The situation in reflexives is, therefore, similar to the case of subject–verb dependencies shown in Example 1.

Numerous studies have found evidence for interference effects in subject–verb dependencies (Dillon, Mishler, Sloggett, & Phillips, 2013; 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, e.g., Dillon et al., 2013, Jäger, Mertzen, Van Dyke, & Vasishth, 2019).

One model that can explain such interference effects is the cue-based retrieval account of Lewis and Vasishth (2005), henceforth LV05. This model is based on the general cognitive architecture ACT-R ("Adaptive Control of Thought-Rational," Anderson et al., 2004; Anderson & Lebiere, 1998). The model was originally implemented as a handcrafted, small-scale 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. The model relies on the core assumptions of ACT-R that retrieving an item from memory is affected by activation decay and similarity-based interference. Using these assumptions, 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: (a) similarity-based interference effects (Dillon et al., 2013; Jäger et al., 2015; Kush & Phillips, 2014; Nicenboim, Logacev, Gattei, & Vasishth, 2016; Nicenboim & Vasishth, 2018; Nicenboim, Vasishth, Engelmann, & Suckow, 2018; Parker & Phillips, 2016, 2017; Patil, Vasishth, & Lewis, 2016; Vasishth, Bruessow, Lewis, & Drenhaus, 2008); (b) the relative roles of predictive processing and memory effects (Boston, Hale, Vasishth, & Kliegl, 2011); (c) impairments in individuals with aphasia (Mätzig, Vasishth, Engelmann, Caplan, & Burchert, 2018; Patil, Hanne, Burchert, De Bleser, & Vasishth, 2016); (d) the interaction between oculomotor control and sentence comprehension (Dotlacil, 2018; Engelmann, Vasishth, Engbert, & Kliegl, 2013); and (e) the effect of working memory capacity differences on underspecification ("good-enough" processing) in sentence comprehension (Engelmann, 2016).

Although the LV05 model has been applied to individual experiments, the empirical coverage of LV05 has never been quantitatively evaluated against a broad range of published benchmark data. 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; Parker, 2019; Rasmussen & Schuler, 2018; 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, Engelmann, and Vasishth (2017). We show that LV05's predictions are restricted to a pattern of two outcomes depending on the experimental manipulation, while the data seem to show evidence for four qualitatively different outcomes instead of two. There may be reasons to suspect that the published patterns of results need to be replicated before they can be regarded as robust (Jäger et al., 2017; Nicenboim et al., 2018; Vasishth, Mertzen, Jäger, & Gelman, 2018). Nevertheless, it is possible that the evidence is reliable; in that case, we would have to conclude that the theory of cue-based retrieval as currently specified by ACT-R and LV05 is wrong, as two out of four patterns seen in the data are not within the model's prediction space. We ask the following question here: Are there systematic, independently motivated assumptions that the ACT-R–based LV05 model is missing that would explain the observed patterns? Toward this end, we develop an extension to the LV05 model to investigate whether relaxing certain assumptions of ACT-R can improve the fit to the observed data. In particular, we question three assumptions in ACT-R that may be oversimplifying the factors that affect retrieval:

1. The assumption that an item's activation is influenced only by its retrieval history and current retrieval cues, while ignoring other factors that may make an item more or less salient, such as its grammatical status or discourse status.
2. The assumption that the strength of the inhibitory interference effect solely depends on the number of distractors rather than on their activation (their availability) in memory.
3. The assumption that each of the cues involved in the retrieval process activates items with one specific feature rather than being associated with multiple features with graded strength.

The effect of salience on activation has been discussed and experimentally investigated but has never been taken into account in any computational model. The second and third assumptions have never been questioned before in experimental studies or computational models of retrieval processes in sentence processing. Using simulations, we explore the predictions of cue-based retrieval that result from relaxing assumptions (1)–(3). The resulting predictions differ from LV05 in several cases and show the existing data under a different light with respect to cue-based retrieval.

Due to the relatively low power of many of the studies that constitute the basis for the model comparisons here, it is not possible in this paper to decide which model is the better one. However, we show that cue-based retrieval could, under specific circumstances, account for certain experimental results that are outside the LV05 prediction space. These circumstances are specified by the model we present here and can guide further research into interference effects in dependency resolution in the search for evidence for or against the ACT-R theory of cue-based retrieval, or indeed any other competing theory.

This paper is accompanied by a web application which can be used to generate predictions with and without our proposed extensions to the model. This application may help in providing insight into the underlying cue-based retrieval mechanisms. The application can be accessed from https://engelmann.shinyapps.io/inter-act/. The underlying code is open source and is available at https://github.com/felixengelmann/inter-act.

Throughout this paper, we use terminology that is specific to cue-based retrieval theory and dependency resolution. In order to help the reader keep track of the terminology, we provide definitions of key terms and concepts in Tables A1 and A2 in the Appendix. Table A1 contains terms related to cue-based retrieval and interference in dependency resolution. Table A2 summarizes new concepts introduced for our extension of cue-based retrieval. Terms that can be looked up in these tables are highlighted in bold when they are introduced in the text.

## 2. A comprehensive quantitative evaluation of the Lewis and 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.

### 2.1. The Jäger et al. (2017) meta-analysis

The meta-analysis had data from 77 experimental comparisons from published eye tracking and self-paced reading studies.[1] 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 terms in boldface which can additionally be looked up in Table A1).

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 stereotypically masculine 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**.[2]

In all the sentences shown in Example 3, the grammatical gender of both target and distractor is manipulated. In this example, the gender match/mismatch of *the surgeon* only refers to its stereotypical gender, which is masculine in English. 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*[3] 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 in the context of this experimental manipulation: 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 (no interference)*

The surgeon$^{\texttt{+MASC}}_{\texttt{+CCOM}}$ who treated Jennifer$^{\texttt{-MASC}}_{\texttt{-CCOM}}$ had pricked himself$\left\{\begin{smallmatrix}\texttt{MASC}\\ \texttt{CCOM}\end{smallmatrix}\right\}$...

b. *Target-match; distractor-match (interference)*

The surgeon$^{\texttt{+MASC}}_{\texttt{+CCOM}}$ who treated Jonathan$^{\texttt{+MASC}}_{\texttt{-CCOM}}$ had pricked himself$\left\{\begin{smallmatrix}\texttt{MASC}\\ \texttt{CCOM}\end{smallmatrix}\right\}$...

c. *Target-mismatch; distractor-mismatch (no interference)*

The surgeon$^{\texttt{-FEM}}_{\texttt{+CCOM}}$ who treated Jonathan$^{\texttt{-FEM}}_{\texttt{-CCOM}}$ had pricked herself$\left\{\begin{smallmatrix}\texttt{FEM}\\ \texttt{CCOM}\end{smallmatrix}\right\}$...

d. *Target-mismatch; distractor-match (interference)*

The surgeon$^{\texttt{-FEM}}_{\texttt{+CCOM}}$ who treated Jennifer$^{\texttt{+FEM}}_{\texttt{-CCOM}}$ had pricked herself$\left\{\begin{smallmatrix}\texttt{FEM}\\ \texttt{CCOM}\end{smallmatrix}\right\}$...

In 3a and 3b, the target matches both cues CCOM and MASC, that is, 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 (because the word *surgeon* has default masculine marking in English) and is thus only a **partial match** for the reflexive. Examples 3c and 3d will, therefore, be referred to as **target-mismatch configurations**.

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. We will, therefore, refer to the distractor-match and distractor-mismatch conditions as **interference** and **no-interference** conditions, respectively. 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 partially matching distractor might either slow down or speed up reading at the critical region, that is, at the reflexive, the reciprocal, or the verb depending on the syntactic construction being considered. Slow-downs and speed-ups are referred to as **inhibitory interference** and **facilitatory interference**, respectively. The term *facilitation* only expresses the fact that a speed-up is observed in behavioral measures, not that processing necessarily becomes easier. 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.

### 2.2. *Predictions of the Lewis and Vasishth (2005) model for target-match and target-mismatch configurations*

See Fig. 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 the match of the item and the higher its **activation level**. The relative activation levels of memory items in ACT-R determine which item will be retrieved. All items available in memory enter into a race at the time of retrieval, such that the one which happens to have the highest activation is retrieved. Thus, only one "winning" item is ever retrieved in any one trial. The higher the activation of the "winning" item, the faster the retrieval time. Each item $i$ has a **base-level activation** $B_i$ that reflects past usage by accounting for all reactivation events ($t_j$ represents the time elapsed since the $j$-th activation) and a time-based decay with rate $d$ (this usually has the default value 0.5 in ACT-R):

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

In the above equation, $\beta_i$ is the resting-state activation for item $i$, and $n$ indexes the number of times that the item $i$ has been retrieved in the past.

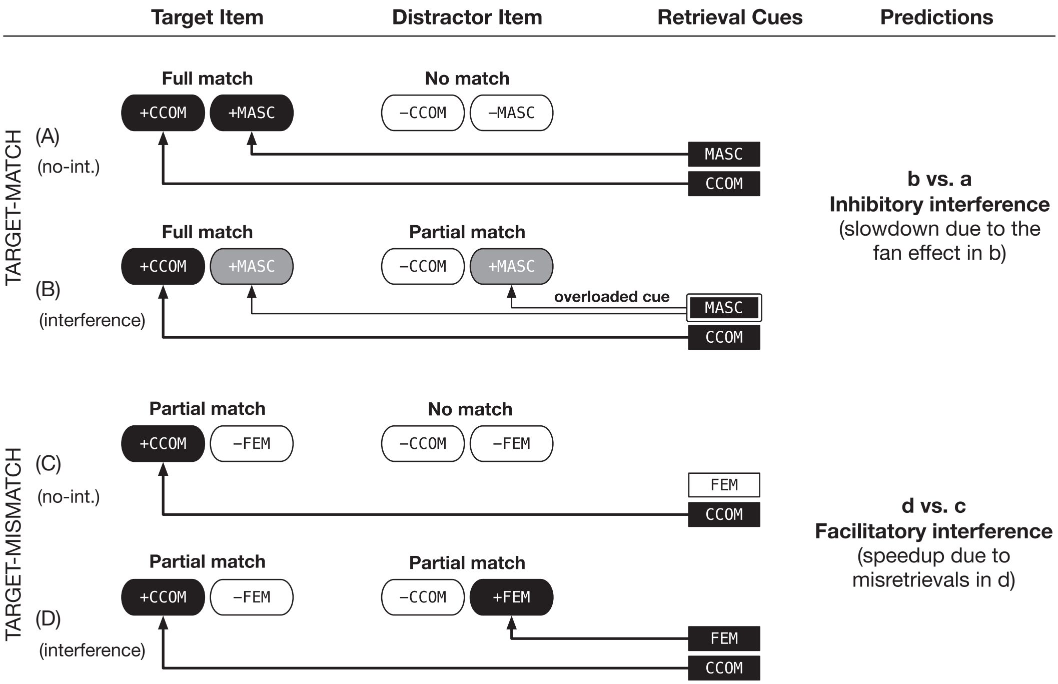



Fig. 1. Spreading activation according to ACT-R/LV05 in the four conditions shown in Example 3. Line weights indicate the amount of spreading activation from a cue to an item. Black oval boxes represent a feature match. Gray oval boxes indicate features matching an "overloaded" cue (MASC in b), and white boxes indicate a mismatch. The figure is by Engelmann and Vasishth (2019); available at https://doi.org/10.6084/m9.figshare.9305456 under a CC-BY4.0 license.

In addition to the base-level activation, **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 $W_j$; see Eqs. 2 and 3. $W_j$ is standardly set to 1/*number of cues*, meaning that all cues are weighted equally. We are adopting this standard assumption throughout this work. (The implications of cue-weighting are discussed in Jäger et al., 2019; Vasishth, Nicenboim, Engelmann, & Burchert, 2019.)

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

The arrows in Fig. 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. In Fig. 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. By contrast, in the interference condition b in Fig. 1 and Example 3, 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). This **cue overload** has the consequence that the activation from this cue is now split between the target and the distractor. This follows from Eq. 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 $S_{ji}$ of this cue with the item $i$. In other words, each competitor reduces the spreading activation to the target by some amount and thus makes it harder to be distinguished from the other items. This is called the **fan effect** (Anderson, 1974). In our example (Fig. 1 and Example 3), the fan effect causes a reduction in 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 Gaussian noise $\epsilon_i$, where $\epsilon_i$ is sampled from a normal distribution with mean 0 and some standard deviation σ (Eq. 4).

$$A_i = B_i + S_i + \epsilon_i, \quad \text{where } \epsilon_i \sim \text{Normal}(0, \sigma) \tag{4}$$

A decrease in activation causes the retrieval time (also called *retrieval latency*) $RT_i$ to increase. As shown in Eq. 5, the retrieval latency of an item is a negative exponential function of its activation at the time of retrieval, where $F$ and $f$ are two scaling parameters—the *latency factor* and the *latency exponent*, respectively.

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

Hence, the similarity in gender between target and distractor in target-match configurations shown in Fig. 1a versus b predicts a slower retrieval latency due to the fan effect, that is, inhibitory interference. At any retrieval event, only the item with the highest activation at that moment is retrieved, and only when its activation is equal or above the retrieval threshold τ. Therefore, the processing time at the word where the retrieval is triggered is dependent only on the time it takes to retrieve the item that happens to have a higher activation, that is, the winner. Due to the Gaussian noise component in Eq. 4, activation fluctuates, such that there is always the possibility—depending on the relative difference in activation between target and distractor—of a **misretrieval**, that is, that the distractor is erroneously retrieved instead of the target. Therefore, because of the increased distractor activation in 1b, there is a higher probability for misretrievals in b compared to a.[4]

In target-mismatch configurations (c and d of Fig. 1 and Ex. 3), the predictions for retrieval latencies are different from those in target-match configurations. 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. Because activation fluctuates due to the random noise component in Eq. 4, when two items receive the same amount of spreading activation from their match with the retrieval cues, the winning item at the time of retrieval is chosen randomly with a probability of around 0.5. Since the winner is always the item with the highest activation—that is, the shortest retrieval latency—at the time of retrieval, this fulfills the conditions of a *race process*. As shown in Fig. 2, in a race process, when the finishing times of two items' retrieval times can be described by distributions that have similar means, the retrieval times of the winner (which can differ from trial to trial) will have a distribution that has a smaller mean than the means of the two items' retrieval time distributions. This is called **statistical facilitation** (Raab, 1962). A race process, therefore, has the effect that, *on average* over multiple trials, the retrieval latency is shorter when the two competing items have similar mean retrieval times than when there is a clear winner due to a bigger difference in retrieval latency as is the case in condition c (e.g., Logacev & Vasishth, 2016).

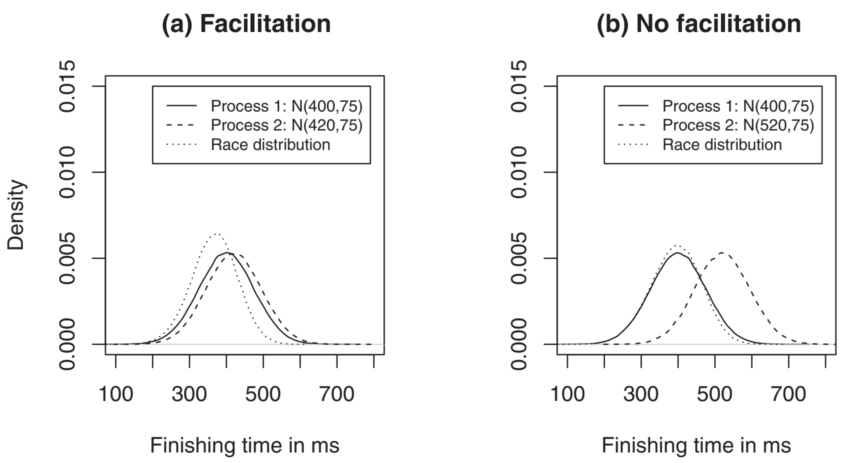



Fig. 2. An illustration of a race process involving two distributions that represent retrieval time distributions of two items. When the two distributions have similar means (a), the distribution of the retrieval times of the winner (which may differ from trial to trial) will have a distribution with a mean that is lower than the mean of the two distributions involved in the race (statistical facilitation). When one distribution has a much smaller mean than the other distribution's mean (b), the distribution of the winner's retrieval times will have the same mean as that of the distribution of the item with the smaller mean.

Because of this statistical facilitation, the prediction for target-mismatch configurations in Fig. 3d versus c is a speed-up on average over multiple trials, that is, facilitatory interference.

## 2.3. Comparison of the LV05 prediction space with the results of the meta-analysis

### 2.3.1. Methods

All simulations part of the present work were run in R (R Core Team, 2016) using the ACT-R equations specified above or—for the extended model—specified on pages 36ff. Additional components of the model that were left out in order not to disrupt the reading flow are provided in Appendix C. All model parameters and their values—if not specified otherwise—are summarized in Appendix Table C1. Simulations were run over a number of trials (and in some cases for a number of parameter values) such that results always represent means. Each iteration generated retrieval time predictions for the four conditions shown in Fig. 1, which were simulated by specifying the respective match between

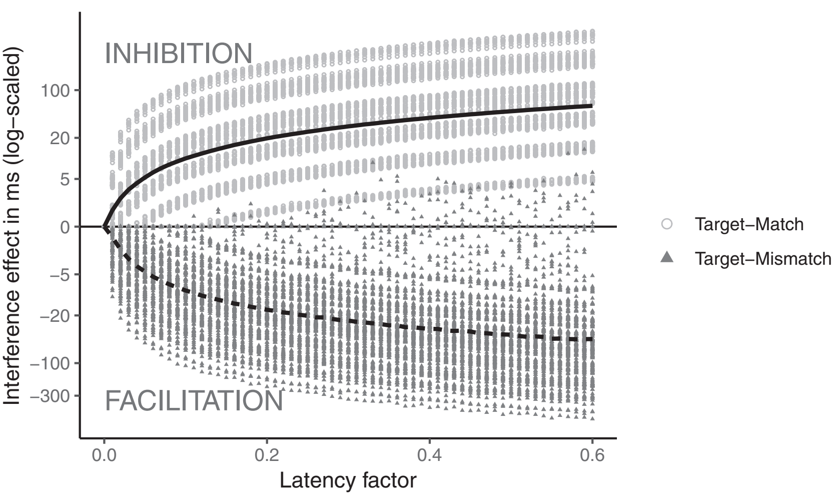



Fig. 3. Prediction space for the interference effect in ACT-R in target-match (circles, solid line) and target-mismatch configurations (triangles, broken line). Interference is plotted in terms of the difference in mean retrieval latencies between the interference (distractor-match) and the no-interference (distractor-mismatch) condition, and as a function of the latency factor $F$. Positive values indicate longer mean retrieval latencies in the interference condition (*inhibitory interference*) due to cue-overload (fan effect) from a partially matching distractor; negative values indicate shorter mean retrieval latencies in the interference condition (*facilitatory interference*) due to retrievals of the partially matching distractor on trials where the distractor is highly activated and hence fast. Each individual data point represents the mean interference effect of 6,000 iterations with 1 out of 10,980 different parameter settings (each in target-match and target-mismatch configurations; that is, there are 21,960 data points plotted in total). Each parameter setting is a combination of the following parameter values: latency factor $F \in \{0, 0.01, \ldots, 0.6\}$, noise parameter $ANS \in \{0.1, 0.2, 0.3\}$, maximum associative strength, $MAS \in \{1, 2, 3, 4\}$ mismatch penalty $MP \in \{0, 1, 2\}$, retrieval threshold $\tau \in \{-2, -1.5, \ldots, 0\}$.

cues and target and distractor respectively as follows: At retrieval, two memory items were available and two retrieval cues were specified. The first (structural) cue was matched by one memory item in all conditions, which distinguished this item as the target.[5] The second cue was matched by the target in conditions a and b (target-match) and by the distractor in conditions b and d (distractor-match). The predicted interference effect was determined for target-match and target-mismatch configurations separately by subtracting the retrieval latency in the distractor-mismatch condition (no interference) from that of the distractor-match condition (interference).

### 2.3.2. Results

Fig. 3 shows the range of possible predictions for the interference effect in target-match and target-mismatch configurations based on the four conditions shown in Fig. 1. The simulations covered range of values for the most relevant ACT-R parameters (see figure caption), which were chosen such that the simulated parameter space included all values commonly used in ACT-R simulations. Values above zero indicate *inhibitory* interference (slow-down) and values below zero indicate *facilitatory* interference (speed-up). Along the $x$-axis of Fig. 3, increasing values of the latency factor $F$ are plotted, 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 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).[6]

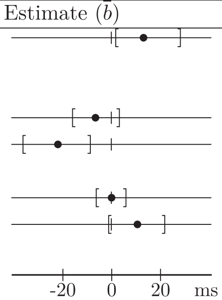
How well do these predictions fare compared to the evidence published in the literature? It turns out that the answer is: not very well. A comprehensive systematic review and meta-analysis of reading studies on interference by Jäger et al. (2017) provides a basis for comparing model predictions with available data. This meta-analysis took into account 77 published experimental comparisons that investigated target-match and target-mismatch configurations for three dependency types. Table 1 summarizes the quantitative results of the meta-analysis. The table shows the mean effect estimates and 95% credible intervals, which mark the uncertainty of the estimates.[7]

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 and usually show different patterns. While agreement studies have focused on effects of number attraction, non-agreement studies investigated interference effects involving other semantic and syntactic cues. Reflexive-antecedent and reciprocal-antecedent dependencies were treated as one category in the meta-analysis because both follow a similar 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

Table 1
Results of the Jäger et al. (2017) meta-analysis showing mean effect estimates $\bar{b}$ with Bayesian 95% credible intervals in the Estimates column. The range specified by a 95% credible interval contains the true value of the estimated parameter with 95% probability, given the model and the data. 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 the extensions item prominence (IP) and multi-associative cues (MAC)

| Dependency | Target | Estimate ($\bar{b}$) | LV05 | +IP | +MAC |
|---|---|---|---|---|---|
| Subject-verb non-agreement | Match | | ✓ | | |
| Subject-verb agreement | Match | | ✗ | ✓ | |
| | Mismatch | | ✓ | | |
| Reflexives/ Reciprocals | Match | | ✗ | ✓ | |
| | Mismatch | | ✗ | | ✓ |




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, 2019; Nicenboim et al., 2018; Vasishth 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). Furthermore, we have recently established in a relatively large-sample (181 participants) eye tracking experiment (Jäger et al., 2019) that in total fixation time, target-mismatch configurations in English reflexives show facilitation effects, as predicted by the ACT-R model. Compare this to one of the studies in the meta-analysis (Dillon et al., 2013), which had a relatively small sample size (40 participants) and found no evidence for facilitatory interference in the target-mismatch reflexive construction.

As discussed in Jäger et al. (2017), one important observation here is that in both target-match and target-mismatch configurations, the individual results of different studies show a considerable range of variability, ranging from facilitatory to inhibitory interference. In the remainder of this article, we explore to what extent an extension of LV05 with independently motivated assumptions can explain the observed variability. We do this in two parts: We first look at the principal consequences of taking into account item prominence, that is, the strength of the distractor's representation in memory relative to the target's, and then explore possible cases and consequences of multi-associative cues.

In both sections, we compare empirical evidence with the prediction space of the revised model that we present. By accounting for item prominence and cue associations on the level of individual studies, the revised model is able to explain some of the facilitatory effects in subject–verb agreement target-match configurations and inhibitory effects in reflexive/reciprocal dependency target-mismatch configurations (as indicated in columns 6 and 7 in Table 1). The apparent absence of a clear effect in the results of the meta-analysis for reflexive/reciprocal dependency target-match configurations can be explained by a mixture of inhibitory and facilitatory effects predicted by the revised model in a principled way as a result of different levels of distractor prominence in individual studies. We then spell out how our revisions to the model are implemented and, finally, present quantitative simulations of the individual studies included in the Jäger et al. (2017) meta-analysis, comparing the estimates from the empirical data with the results of both LV05 and the revised model.

## 3. An extension of the LV05 model

We reconsider three assumptions in the ACT-R–based, cue-based retrieval model that constitutes the basis of the LV05 predictions. These are the following:

1. The base-level activation of items in memory is a function only of decay and reactivation through study-relevant retrieval events. Other influences (discussed in detail below) are usually ignored.
2. The fan effect (the inhibitory interference effect caused by cue overload) is a function of the number of items that match a specific retrieval cue, independent of their activation.
3. The associative strength between a retrieval cue and a memory item is based on a binary (match/mismatch) one-to-one mapping between the cue and a feature value.

These assumptions are, in fact, oversimplifications that do not accurately reflect general aspects of cognition. In particular, considering that the memory activation of an item represents its strength of representation or its accessibility, it should (a) affect the strength of interference, and (b) take into account more aspects of the linguistic context than only the retrieval event that is relevant in a particular experiment. Furthermore, given that cognitive associations between contextual cues and certain representation are the result of associative learning through experience, one should account for the fact that these associations can be graded and multi-associative in nature and are not necessarily strictly categorical. Motivated by these considerations, the revised assumptions we propose are as follows:

1′. The base-level activation of items in memory (i.e., accessibility) is affected by—in addition to recency—their **prominence** in the current context, that is, their relevance/salience in terms of syntactic relations in a sentence or information-structural and discourse properties.

2′. The strength of any interference effect—including the fan effect—is not simply a function of the presence versus the absence of a distractor, but changes as a function of the distractor's activation in memory relative to the target.
3′. The associative strength between a retrieval cue and a memory item can be the result of multiple cues being associated with multiple features at variable degrees. **Cue-feature associations** are based on associative learning through language experience.

In the following two sections, we show how the revised assumptions change the prediction space of LV05 and how this compares to the empirical evidence. We begin with an investigation of the way that different levels of distractor prominence can change the predictions (revised assumption 1′), assuming that relative activation affects the strength of the fan effect (revised assumption 2′).

### *3.1. Item prominence*

The activation of a noun in memory prior to being retrieved (its base-level activation) is usually considered as a function of the time since it is encountered in the sentence (cf. Eq. 1). However, whether it was introduced as a subject or an object might change the way the noun is maintained in memory. Similarly, if a noun has been introduced in a context sentence previously, it may affect its memory representation. Indeed, independent evidence shows that the accessibility of a noun phrase is increased in prominent grammatical positions or through increased discourse saliency, such as being the discourse topic (Ariel, 1990; Arnold, 2007; Brennan, 1995; Chafe, 1976; Du Bois, 2003; Grosz, Weinstein, & Joshi, 1995; Keenan & Comrie, 1977). It is plausible that items which have a high prominence by virtue of their grammatical or discourse status are retrieved intermittently or maintained with high activation in memory. This implies that their base-level activation is higher than that of less prominent items due to reactivation boosts and reduced decay. More prominent items would thus have an elevated activation level prior to retrieval and will therefore—other things being equal—be retrieved with higher probability and lower latency than items with lower prominence.

In the same way, prominence could include other factors that we do not consider here: For example, thematic role (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 an item's activation in memory. We focus here on the effects of grammatical position and discourse status, which have been discussed in the literature on memory interference in dependency resolution (Cunnings & Felser, 2013; Patil, Vasishth, & Lewis, 2016; Sturt, 2003; Van Dyke & McElree, 2011).

In the model, instead of modeling each additional hypothesized retrieval or reactivation event, we simply add a term to the base-level activation that is a function of the grammatical role or discourse status of the memory item. Because of our revised assumption 2′, according to which the magnitude of the interference caused by a distractor in the model depends on its activation relative to the target, a sentence containing a high-prominence distractor should show a different interference effect than a sentence with a low-

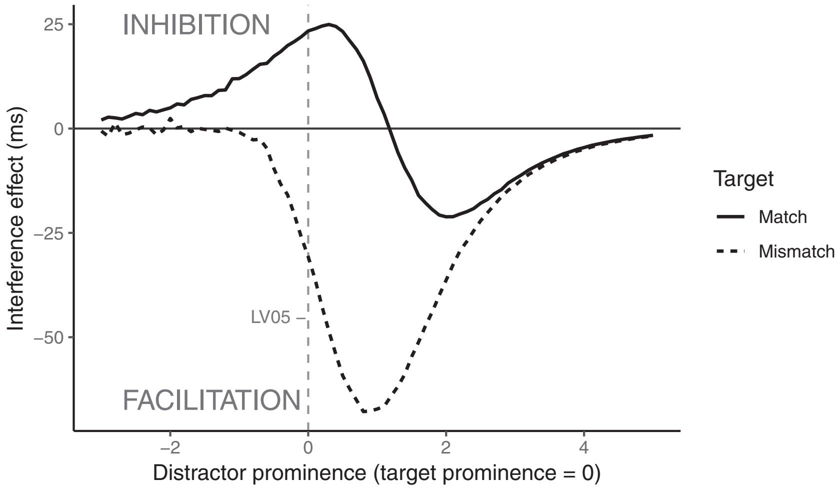



Fig. 4. Predicted target-match and target-mismatch interference effects (distractor-match minus distractor-mismatch) as a function of distractor prominence $p_{dstr} \in \{-3, -2.9, \ldots, 5\}$ when target prominence is zero (mean of 10,000 iterations with parameters $F$ = 0.2, $ANS$ = 0.2, $MAS$ = 2, $MP$ = 0). Positive values indicate longer mean retrieval latencies (inhibition) in the interference condition due to cue-overload (fan effect). Negative values indicate shorter mean retrieval latencies (facilitation) in the interference condition due to retrievals of the distractor on trials where the distractor is highly activated and hence fast. The points where the vertical line intersects with the curves represent standard LV05 predictions.

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.

The full details on the implementation will be presented in the section beginning on page 36. Here, we already show the results of simulations with the extended model in order to illustrate the general predictions as a function of prominence. Fig. 4 shows the interference effect predicted by our model as a function of the prominence of the distractor $p_{dstr}$ (in terms of its base-level activation) with respect to the prominence of the target, which stays constant at zero.

Overall, in target-mismatch configurations, there is only facilitation, which increases with higher distractor prominence. In target-match, low values of distractor prominence produce inhibition, while high values produce facilitation. In order to understand the causes that drive the behavior of the model, it is important to be clear about how the data in Fig. 4 were generated: (a) The interference effect shown in Fig. 4 is the latency difference between the interference condition (when the distractor matches one of the retrieval cues) and the no-interference condition (when the distractor does not match the retrieval cues). The interference effect, therefore, reflects how distractor prominence affects both these conditions. (b) The effects shown are computed from mean retrieval latencies per condition across multiple trials. (c) The latency values in each trial are a function of the activation value of only the most activated (hence, retrieved) item, which can be either the target or the distractor. Hence, the mean latency in each condition reflects a mix of target and distractor activation values. (d) The distractor in the no-interference condition is always less activated than the distractor in the interference condition, because the latter matches one of the retrieval cues and, therefore, receives spreading activation.

Inhibitory interference is a consequence of the fan effect, which only occurs in target-match configurations, where activation that is associated with the non-structural cue (gender or number) is split between the target and the distractor in the interference condition. When distractor activation is very low, the fan effect is weak. When the prominence of target and distractor is roughly the same, the distractor is strong enough to cause a fan effect but still weak enough that the target is retrieved most of the time in both conditions (because it matches the structural cue which the distractor does not). Thus, the fan effect on the target causes inhibitory interference (the interference condition is processed slower than the non-interference condition).

When the activation level of target and distractor is equal in the interference condition after taking into account prominence and cue match, this leads to *statistical facilitation* due to a race process between two similarly activated items (Logacev & Vasishth, 2016; Raab, 1962). In target-match, this is the case at a prominence value of about 0.9 when distractor prominence compensates for the activation difference between target and distractor, which is because the distractor only matches one of the two cues. Here, the inhibitory interference effect starts to decrease because it is counteracted by the statistical facilitation. In the target-mismatch configuration, statistical facilitation occurs when the difference in prominence of target and distractor is equal, which is equivalent to the predictions of the original LV05 model, here represented by the vertical line in Fig. 4.

The rest of the pattern in Fig. 4 is, however, independent of statistical facilitation. In both target-match and target-mismatch configurations, high values of distractor prominence lead to increased facilitation effects. The reason is that, when the distractor activation is above that of the target, the race in the interference condition is won by the distractor most of the time. In the no-interference condition, however, at first, the target is still winning most of the time, since it matches more retrieval cues. Therefore, the average processing time in the interference condition mainly reflects the increasing activation of the distractor, while in the no-interference condition, it reflects the static activation of the target. While this is the case, increasing the distractor prominence will increase the difference between both conditions and, hence, increase the facilitatory interference effect. For both target-match and target-mismatch, this dynamic stops as soon as the distractor is winning most of the time in both the interference and no-interference conditions. As activation goes to infinity with increasing prominence, retrieval latencies in interference and no-interference conditions asymptote to zero (see ACT-R latency equation 5). Therefore, the interference effect—which is the difference between interference and no-interference—in both target-match and target-mismatch configurations also asymptotes toward zero.

For a detailed explanation of the mechanisms behind the pattern shown in Fig. 4, see the Supplementary Material file "Sup Mat Prominence.pdf" provided with this article.

How do these predictions match up with the data? In the literature on target-match interference configurations with high-prominence distractors, there is some evidence for both (A) inhibitory effects as well as (B) facilitatory effects. For the remainder of this section, we summarize this evidence in comparison with the predictions shown in Fig. 4.

*A: Target-match inhibition*: In an eye tracking and a speed-accuracy trade-off 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 <u>**the witness/the motion** was inappropriate</u> realized that the attorney in the case compromised
b. The judge who <u>had rejected **the witness/the motion**</u> realized that the attorney in the case compromised.

Patil, Vasishth, and Lewis (2016) found an interference effect at the reflexive in an eye tracking experiment using sentences as in 5 with the distractor *Fred* in subject position, which was a modification of Sturt (2003) (shown 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 the manipulation of Van Dyke and McElree (2011), a prominent distractor (in subject position) in a target-match configuration caused inhibitory interference while a non-prominent distractor (in object position) did not. As shown in Fig. 4, our prominence model predicts that the inhibitory effect in target-match configurations increases with higher distractor prominence up to a certain point. Hence, the model predictions fit the data here if we assume that a distractor in subject position causes a medium increase in prominence, as opposed to a major increase, which would lead to facilitation. More specifically, the prediction is correct under the assumption that, in the interference condition, the distractor activation is still lower than the target activation (see region between points a and b in Fig. S1.1 in the Supplementary Material file "Sup Mat Prominence.pdf"). As discussed below, a distractor being the discourse topic seems to cause facilitatory interference in target-match. Therefore, in our model, subject position would be located at a medium value and discourse topic at a high value along the continuum of distractor prominence.

In a reflexive-antecedent study in Mandarin Chinese, Jäger et al. (2015) found a similar difference in target-match configurations between their Experiment 1, where a distractor was present in the sentence, and their 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 while reading/comprehending the target sentence would make them more prominent in memory, that is, increase their activation, which would amplify the interference effect, again as shown in Fig. 4.

*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* had been made the discourse topic using a context sentence. Cunnings and Felser used sentences such as Example 7 and the baseline condition shown in Example 8, where the distractor noun phrase was introduced in a context sentence and was co-referred to in the target sentence through the pronoun *he*. The authors hypothesized that the distractor was more prominently encoded due to reactivation at the anaphor, and that this may have increased the probability of observing an interference effect at the reflexive (pp. 212–213).

(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.

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

A distractor that is a discourse topic and is also in subject position would arguably be more prominent than if it were just in subject position but not a discourse topic. The qualitative difference of the target-match effects described above is thus predicted by our model. As explained above, a distractor that is more activated than the target causes an increasing number of retrievals of the highly activated distractor, which has a faster retrieval latency and, therefore, yields a facilitatory interference effect on average when the speed-up is strong enough to counteract the fan effect.

In summary, the integration of prominence in the form of base-level activation (assumption $1'$) and the fan effect being a function of distractor activation (assumption $2'$) can explain *inhibitory* interference effects 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 *facilitatory* interference effects 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 predicts neither the facilitatory interference effects in target-match configurations nor the systematic absence of an effect under certain conditions. Earlier, in Table 1, we indicated 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 of low, medium, or high prominence values on the continuum shown in Fig. 4.

Next, we investigate the prediction space of LV05 under assumption 3′ that retrieval cues can be associated with multiple features to varying degrees.

### 3.2. Multi-associative cues

The noun *surgeon* in the sentence "the surgeon who treated Jennifer had pricked herself" receives increased activation in memory when retrieval is triggered at *himself* because it matches the syntactic cue CCOM. The distractor *Jennifer* receives the same amount of activation through its match with the gender cue FEM. As depicted in Fig. 1d, this leads to the situation with two similarly activated items, but no fan effect because their features do not overlap—each item is associated with a different cue. This leads to statistical facilitation in the way explained earlier.

In ACT-R models, a match between a cue and a feature is binary and categorical: A feature and a cue can only match or not match (there is no gradation); and 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. Such categorical one-to-one relations are, of course, a simplification made for modeling, but they are not well motivated when we accept that language acquisition is essentially a gradual process of learning the mapping from form to meaning on the basis of contextual cues (see, e.g., Bybee, 2006; Langacker, 1987; Tomasello, 2003). The strength and distinctiveness of representational associations are thus dependent on *similarities* with other associations. In that sense, retrieval cues represent abstract knowledge about the features that successfully identify the correct retrieval target, as derived from experience with a certain dependency context. Hence, cue-feature associations evolve as graded associations between a retrieval context and any features of the correct target resulting from a process of learning relevant

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

| Context | Target Features | Form |
|---|---|---|
| EN reflexive | $\begin{Bmatrix}\text{+MASC}\\\text{+CCOM}\end{Bmatrix}$ | himself |
| | $\begin{Bmatrix}\text{+FEM}\\\text{+CCOM}\end{Bmatrix}$ | herself |
| | $\begin{Bmatrix}\text{+NEUT}\\\text{+CCOM}\end{Bmatrix}$ | itself |
| | $\begin{Bmatrix}\text{+PLUR}\\\text{+CCOM}\end{Bmatrix}$ | themselves |
| EN reciprocal | $\begin{Bmatrix}\text{+PLUR}\\\text{+CCOM}\end{Bmatrix}$ | each other |
| CN reflexive | $\begin{Bmatrix}\text{+ANIM}\\\text{+CCOM}\end{Bmatrix}$ | ziji |

discriminations between features. As a result, it is possible that, in certain situations, a cue can be associated with *multiple* feature values to varying degrees.

Now, there is no specific reason why, in the sentence "the surgeon who treated Jennifer had pricked herself," the syntactic cue would be associated with gender features or vice versa. But consider, for example, the sentence "the nurse who cared for the children had pricked each other." The relevant retrieval cues for the reciprocal *each other* are CCOM and PLURAL. The cue CCOM is matched by the syntactically correct target *the nurse* and PLURAL is matched by the distractor *the children*. The difference between reciprocals, such as *each other*, and reflexives, such as *herself*, is that the correct target in a reciprocal context *always* exhibits the features +PLUR and +CCOM, while there are several possible forms in a reflexive context, for example, *himself*, *herself*, *itself*, and *themselves*, which all trigger different combinations of the syntactic cue with gender and number cues, as listed in Table 2.

The CCOM cue in reflexive contexts would, therefore, not be strongly associated with, for example, the +FEM feature because this would activate the wrong items whenever the form of the reflexive is not *herself* but *himself* or *themselves*, etc. Therefore, the reflexive context requires syntactic, gender, and number features to be discriminated. In reciprocal contexts, however, the correct target has to be plural. In this case, if the CCOM cue were to be associated with the +PLURAL feature, it would always activate the correct target. Because +CCOM and +PLURAL *co-occur* frequently in similar contexts for reciprocal-antecedent dependencies, a strong discrimination is less required than in reflexive contexts. Instead, it might even be more efficient to also activate plural items with the CCOM cue and vice versa. This reasoning builds on the ideas of classical conditioning (Rescorla & Wagner, 1972), where two stimuli that require similar responses in similar contexts become less discriminated than when they elicit different responses. As a consequence, the cues CCOM and PLURAL in reciprocal-antecedent dependencies would be less **discriminative** than the cues in reflexive-antecedent dependencies and

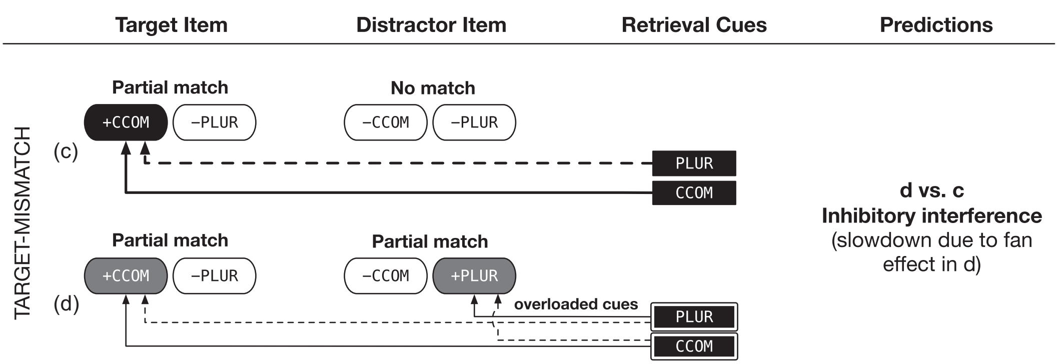



Fig. 5. Spreading activation in distractor-match (c) and distractor-mismatch (d) conditions in target-mismatch configurations when cues are cross-associated. Line weight and box shading indicate the amount of spreading activation added to an item due to a feature match. Dashed lines represent spreading activation to a cross-associated feature.

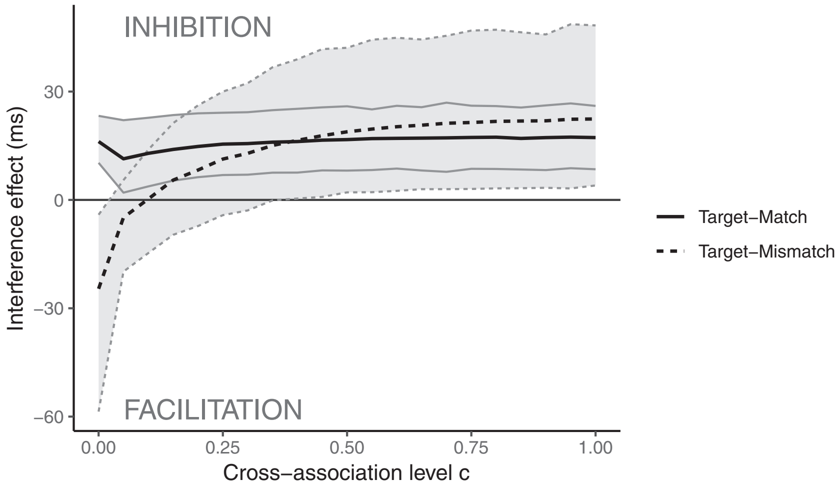



Fig. 6. Predicted target-match and target-mismatch interference effects (distractor-match minus distractor-mismatch) as a function of the cross-association level *c*. Lines and shaded area show mean and range of the effect, respectively, for parameter values of the latency factor $F \in \{0.2, 0.3, 0.4\}$ and distractor prominence $p_{dstr} \in \{-0.5, 0, 0.5\}$, running 5,000 iterations each; other parameters were fixed as $ANS = 0.2$, $MAS = 2$, $MP = 0$. Positive values indicate longer mean retrieval latencies (inhibition) in the interference condition due to cue-overload (fan effect). Negative values indicate shorter mean retrieval latencies (facilitation) in the interference condition due to misretrievals of the distractor.

would, therefore, both be associated to some degree with both the features +CCOM and +PLUR. We will say that, in this situation, two cues are **cross-associated** due to **feature co-occurrence**. A similar situation arises for the Chinese reflexive *ziji* (also shown in Table 2), which requires an animate and c-commanding target. Thus, in the case of *ziji*, CCOM would be cross-associated with ANIM.

For an illustration of the predictions that would arise from cross-associated cues, consider Fig. 5. The figure shows the no-interference (c) and interference (d) conditions in *target-mismatch* configurations when cues are cross-associated in contrast to Fig. 1, where no cross-association was present. 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; that is, they are overloaded. This leads to a similar situation as in target-match configurations shown earlier in condition b of Fig. 1: As spreading activation is shared between target and distractor, inhibitory interference, that is, 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 versus c.[8]

In order to explore the quantitative consequences for the predicted interference effect, we implemented cross-associated cues in an extension of LV05 and ran simulations with a range of values for the *cross-association level*. As the results in Fig. 6 clearly show, an increasing cross-association level causes an inhibitory fan effect in target-mismatch configurations that eliminates the facilitatory effect.

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 c = 1 means that their corresponding features are treated as functionally identical; that is, each cue activates both features.

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

$$c_{kl}(Context) = \frac{\sum [k \wedge lContext]}{\sum [k \vee lContext]} \tag{6}$$

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, for example, that the cross-association levels for the examples in Table 2 are for reflexives $c_{\mathrm{CCOM,MASC}}(\textit{refl-EN}) = 1/4 = 0.25$, for reciprocals $c_{\mathrm{CCOM,PLUR}}(\textit{reci-EN}) = 1/1 = 1.0$, and for *ziji* $c_{\mathrm{CCOM,PLUR}}(\textit{ziji}) = 1/1 = 1.0$. 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 one hand and English reciprocals or *ziji* on the other. What this calculation suggests is that, when processing English reflexives, more distinct cue representations are used due to a greater variety of feature combinations than for reciprocals or *ziji*.

In summary, the theory of multi-associative cues predicts that a cue could 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. Table 1 shows inhibitory instead of facilitatory interference in target-mismatch configurations. This has been found, for instance, in some studies on reflexives and reciprocals and can be explained neither 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. Our approach of using 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,[9] 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.

### 3.3. 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 such that the similarity between features is also used in the computation of the fan effect.

The general idea of our extension is that each item's prominence as well as specific cue-feature associations are reflected in the associative strength $S_{ji}$ between a cue $j$ and an item $i$, which in turn affects the activation $A_i$ of that item. In other words, the associative strength that a memory item has with a specific cue reflects the prominence status of all memory items and the relative associations of that cue with *all features of all memory items*. Therefore, the two mechanisms' item prominence and multi-associative cues are merely two aspects of one broader mechanism, namely the association of the available retrieval cues with specific memory items. In order to incorporate prominence and multi-associative cues, we redefine the associative strength $S_{ji}$. Recall from Eq. 4 that, given a set of retrieval cues ($Cues = \{q1, \ldots, qJ\}$), the activation $A_i$ of an item $i$ is a function of spreading activation $S_i$:

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

For each cue $j$, the standard ACT-R calculation of $S_{ji}$ is based on its 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). 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 $j$:[10]

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

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** $Q_{ji}$ (which will be defined further below) of item $i$ with cue $j$ in proportion to the match quality $Q_{jv}$ of all active memory items $v$ with $j$:

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

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

### 3.4. Multi-associative cues

We assume that a cue can have variable discrimination, that is, it can be associated with multiple features to different degrees. The associative strength between a cue $j$ and a feature $k$ is given by $M_{jk}$, which takes values between 0 (not associated) and 1 (maximally associated). 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 $i$.

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

As shown in Fig. 6, 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 that an item $i$ has feature f1 but not feature f2, and a distractor item i' has feature f2 but not f1 (see Fig. 7). Assume also that the retrieval cue q1 matches f1, and cue q2 matches f2. Retrieval is triggered using the two cues q1 and q2. This is the typical target-mismatch/distractor-match scenario discussed earlier.

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 q1 depends on the probability of item $i$ given cue q1:

$$P(i|q1) = \frac{Q_{q1,i}}{\sum\limits_{v \in Items} Q_{q1,v}} \tag{11}$$

The numerator is computed as follows. Since only feature f1 matches cue q1 in item $i$, we have:

$$Q_{q1,i} = \sum_{k \in K_i} M_{q1,k} = M_{q1,f1} = 1 \tag{12}$$

The denominator, $\sum\limits_{v \in Items} Q_{q1,v}$, also has value 1 because it is the sum of the match of cue q1 to item $i$ (which is 1) and to item $i$' (which is 0):

$$\sum_{v \in Items} Q_{q1,v} = Q_{q1,i} + Q_{q1,i'} = 1 + 0 \tag{13}$$

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

$$P(i|j) = P(i|q1) = \frac{Q_{q1,i}}{\sum\limits_{v \in Items} Q_{q1,v}} = \frac{1}{1} = 1 \tag{14}$$

This implies that the spreading activation from cue q1 to item $i$ is:

$$\begin{aligned} S_{q1,i} &= MAS + \ln[P(i \\ &= MAS + \ln[1] = MAS \end{aligned} \tag{15}$$

As no other cue matches item $i$, $S_{q1,i}$ equals the total amount of spreading activation $S_i$ that item $i$ receives:

$$S_i = S_{q1,i} = MAS \tag{16}$$

Thus, there is no penalty to the activation of item $\underline{i}$ 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 q1, but also from q2, which spreads 50% of its activation to feature f1. The distractor $i$' similarly gets activation not only from q2, which fully matches f2, but also from q1, which spreads 50% of its activation to feature f2. Graphically, this corresponds to the following scenario (Fig. 8).

Now, $P(i|j)$ is not 1 but 1/1.5 or 2/3.

$$P(i|q1) = \frac{Q_{q1,i}}{\sum_{v \in Items} Q_{q1,v}} = \frac{1}{1.5} = \frac{2}{3} \tag{17}$$

This is because $Q_{q1,i} = 1$ as in the previous calculation, but the denominator is the sum of the match of cue q1 to item i (a match of 1) as well as the match of cue q1 to item *i'* (a match of 0.5).

$$\sum_{v \in Items} Q_{q1,v} = Q_{q1,i} + Q_{q1,i'} = 1 + 0.5 = 1.5 \tag{18}$$

We then use $P(i|j)$ to calculate the spreading activation $S_{q1,i}$ from cue q1 to item *i*. In contrast to the scenario above without cross association, the amount of activation spread $S_{q1,i}$ is now smaller than *MAS*:

$$S_{q1,i} = MAS + \ln\left[\frac{2}{3}\right] = MAS + [-0.41] = MAS - 0.41 \tag{19}$$

Next, the calculation for item *i* and cue q2 is:

$$P(i|q2) = \frac{Q_{q2,i}}{\sum_{v \in Items} Q_{q2,v}} = \frac{0.5}{1.5} = \frac{1}{3} \tag{20}$$

Here, $Q_{q2,i} = 0.5$ because of the cross-association of 0.5 of cue q2 with the feature f1. The denominator is the sum of the match of cue q2 to item *i* (a match of 0.5) as well as the match of cue q2 to item *i'* (a match of 1).

$$\sum_{v \in Items} Q_{q2,v} = Q_{q2,i} + Q_{q2,i'} = 0.5 + 1 = 1.5 \tag{21}$$

We now use $P(i|q2)$ to calculate the spreading activation $S_{q2,i}$ that item *i* receives from cue *q2*. Similar to $S_{q1,i}$, $S_{q2,i}$ will also be smaller than *MAS*:

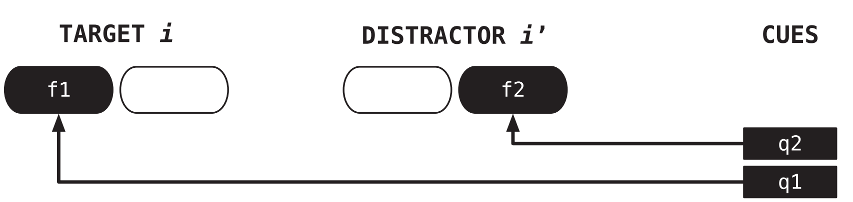



Fig. 7. Standard target-mismatch/distractor-match condition without cross-associated cues.

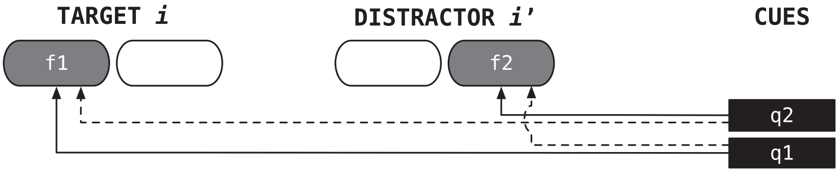



Fig. 8. Target-mismatch/distractor-match condition when cues are cross-associated.

$$S_{q2,i} = MAS + \ln\left[\frac{1}{3}\right] = MAS + [-1.1] = MAS - 1.1 \tag{22}$$

Having computed $S_{q1,i}$ and $S_{q2,i}$, 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):

$$\begin{aligned} S_i &= \sum_{j \in Cues} W_j S_{ji} \\ &= \frac{1}{2} S_{q1,i} + \frac{1}{2} S_{q2,i} \\ &= \frac{1}{2}\left(MAS + \ln\left[\frac{2}{3}\right]\right) + \frac{1}{2}\left(MAS + \ln\left[\frac{1}{3}\right]\right) \\ &= MAS + \frac{1}{2}\left(\ln\left[\frac{2}{3}\right] + \ln\left[\frac{1}{3}\right]\right) \\ &= MAS - 0.75 \end{aligned} \tag{23}$$

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 the cross-association level is sufficiently high.[11]

### 3.5. 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 Eq. 1 for $B_i$ is changed to:

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

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 Eq. 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, $\tau$ 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 discriminative* (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.

Fig. 4 shown earlier illustrates the relationship between distractor prominence and the interference effect as predicted by the extended model, assuming that target prominence is a fixed 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 discrimination (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 discrimination of cues and the prominence of the item itself and of all of its competitors affect the fan effect, that is, 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.

## 4. A simulation of all the studies in the Jäger et al. (2017) meta-analysis

In this section, we compare the specific predictions of LV05 and an extended model with item prominence and multi-associative cues for the experiments in the Jäger et al. (2017) meta-analysis. We make the following assumptions with regard to the relation

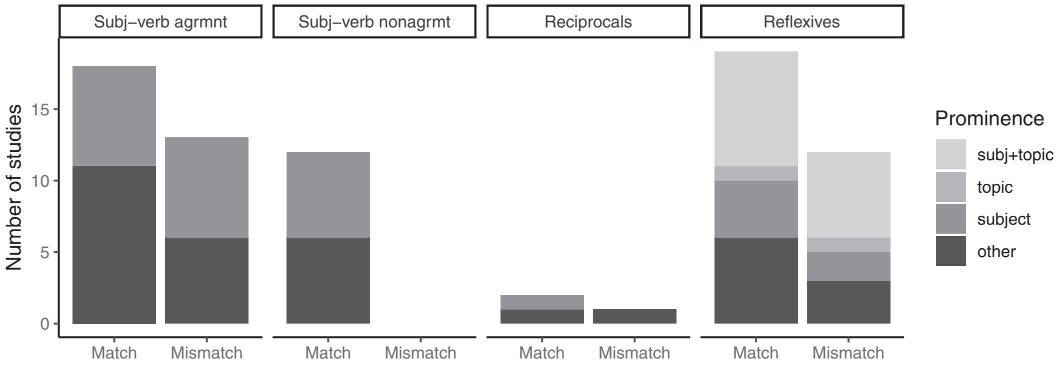



Fig. 9. Number of studies included in the Jäger et al. (2017) meta-analysis and in the simulations, grouped by dependency type and distractor prominence status (studies are listed in Table B1 in the Appendix).

between a specific experiment and the extended model settings for prominence and cue-feature associations:

- Being a sentential subject and being mentioned in a context sentence (discourse topic) both increase the prominence—and hence the base-level activation—of a target or distractor compared to being an object and not the discourse topic. The combination of both (subject and topic) has the highest prominence status.
- The cue-feature cross-association level is raised only for dependency contexts where the cues can be assumed to have low discrimination due to feature co-occurrence. These contexts are experiments involving reciprocals and those involving the Chinese reflexive *ziji*.

The simulations presented here can be reproduced using the accompanying web application at https://engelmann.shinyapps.io/inter-act/. The model code is available on GitHub at https://github.com/felixengelmann/inter-act.

### *4.1. Data*

We included all studies that were part of the meta-analysis in Jäger et al. (2017). Table B1 in the Appendix lists all included studies with their dependency types and distractor prominence levels. Fig. 9 shows the number of target-match and target-mismatch comparisons for each dependency type and prominence category. At the time of the meta-analysis, no data were available for target-mismatch 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 Section 5. We categorized the experiments into three different prominence relations for the distractor: *subject position*, *discourse topic*, and *other*. Subject position and *discourse topic* are considered high prominence levels, while we do not make any a priori assumptions about which of the two 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 a discourse topic. We expect the prominence in this case—and thus the distractor activation—to be particularly high. As the figure shows, distractors that are a discourse topic, and the combination of discourse topic and subject position has so far only been tested in reflexives.

### 4.2. Method

The extended model as described above was implemented in R (R Core Team, 2016). Prominence and multi-associative cues could be switched off such that the model behavior is then equivalent to LV05. The model was set up as described on page 15 in order to simulate the four conditions shown in Fig. 1, representing retrieval processes in sentences similar to Example 3. Different from the simulations above, multiple distractors were specified for some of the studies that were part of the present simulation. The model was run for 5,000 iterations on each experiment and yielded the mean effect sizes for target-match and target-mismatch configurations, which in each case were determined by subtracting the retrieval latency in the distractor-mismatch condition from that of the distractor-match condition.

#### 4.2.1. Parameter estimation

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. Certain parameters were estimated by running the model iteratively while changing the parameter value within a pre-specified range (details are discussed below). The best value was determined by finding the lowest mean-squared error between the simulated and experimental effects using grid search.

As is common practice in ACT-R modeling, we estimated the latency factor $F \in \{0.1, 0.125, \ldots, 0.25\}$ (see Eq. 5) for each experiment in both models to scale the numerical results into a range that is comparable with the data. In the extended model, the distractor prominence parameter $p_{dstr}$ was estimated across experiments for each of three prominence categories within dependency types: *low* (neither subject nor topic), *medium* (subject *or* topic), and *high* (subject *and* topic). For each of these categories, $p_{dstr}$ was restricted to a certain range that was determined according to the pattern in Fig. 4 as follows: Medium prominence was constrained to be close to the target prominence ($p_{trgt} = 0$) in the area where the distractor has an influence on the fan effect of the target ($p_{dstr} \in \{-1, -0.9, \ldots, 2\}$); low prominence was constrained to be smaller than the target prominence ($\{-2.5, -2.4, \ldots, 0\}$); and high prominence was bound to values higher than the target prominence and above the point where in Fig. 4 the target-match fan effect begins to disappear ($\{1, 1.1, \ldots, 4\}$).

Thus, the full range of predictions shown in Fig. 4 can be generated theoretically, but the generating process is restricted to specific properties of the distractor. Without restricting the prominence parameter in this way, the model cannot be fit in a meaningful way because some predictions can result from multiple prominence values. This can be seen in Fig. 4 (e.g., the absence of a target-match effect is predicted at very low, very high, and at a medium prominence just over 1). The value ranges were allowed to overlap, however, in order not to pre-impose any assumptions about specific effect sizes on the model. The target, which was a subject in all experiments, was assumed to have equal prominence across experiments. Its prominence value was, therefore, set to 0.

The cross-association level $c$ was estimated only for the two cases we have mentioned above: reciprocals and the Chinese reflexive *ziji*. It was estimated in these cases within $\{0.1, 0.2, \ldots, 1\}$ and 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 interference designs, the target was more distant from the retrieval site than the distractor, and vice versa for proactive interference designs. Hence, interference type affects the model through the memory decay component, which reduces the activation of an item as a function of time.

### 4.3. Results

We ran simulations both with the original LV05 model and with the extended model that included item prominence and multi-associative cues (LV05 + IP + MAC). Because Lewis and Vasishth (2005) speculated that model fit might improve without the decay component of ACT-R,[12] we also ran variants of both models without the decay component.

Table 3 summarizes the fit for all four model configurations in terms of the root-mean-square deviation, averaged within dependency types. Overall, the extended model with IP and MAC 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 and data considered here. We revisit this point in Section 5.

More important than the numerical fit of a computational model with the data, however, is that the model correctly reproduces observed patterns in a principled way. As we saw in Table 1 earlier, certain observed patterns were incompatible with LV05,

Table 3
Root-mean-square deviation between modeling results and observed data, averaged within dependency type and model (best values in bold). The superscript no dec means that the decay parameter is set to 0

| Dependency | LV05 | LV05$^{no\ dec}$ | LV05 + IP + MAC | LV05 + IP + MAC$^{no\ dec}$ |
|---|---|---|---|---|
| Subject–verb agreement | 18.06 | 15.54 | 14.47 | **13.03** |
| Subject–verb non-agreement | 7.04 | 7.85 | **5.04** | 7.96 |
| Reflexives/Reciprocals | 12.40 | 11.68 | 7.46 | **6.3** |

specifically, facilitatory interference in subject–verb agreement target-match configurations, inhibitory interference in reflexive/reciprocal target-mismatch, and the absence of an effect in reflexive/reciprocal target-match. The content of Table 1 is repeated here graphically in Fig. 10. The figure shows the estimates for the mean effect along with 95% credible intervals, as well as the average simulated effects of both models for the same three dependency categories as in Table 1. A qualitative improvement can be seen in reflexive/reciprocal dependencies. The extended model's results are within the 95% credible intervals of the data estimates, showing that LV05 + IP + MAC can potentially explain why no effect was found in target-match and inhibition was found in target-mismatch. In terms of the facilitatory effect in subject–verb agreement target-match configurations, the extended model does not show a qualitative difference to LV05 but merely shows a smaller effect that is closer to the mean estimated from the data.

The crucial difference between the two models is that the LV05 cannot explain the data that show facilitation in target-match or inhibition in target-mismatch configurations. The extended model, however, can account for these patterns when they can be explained by distractor prominence and cross-associated cues.

This becomes more apparent when presenting the means by distractor prominence levels as in Fig. 11. Here, the simulated means are compared to the *sample means* estimated from

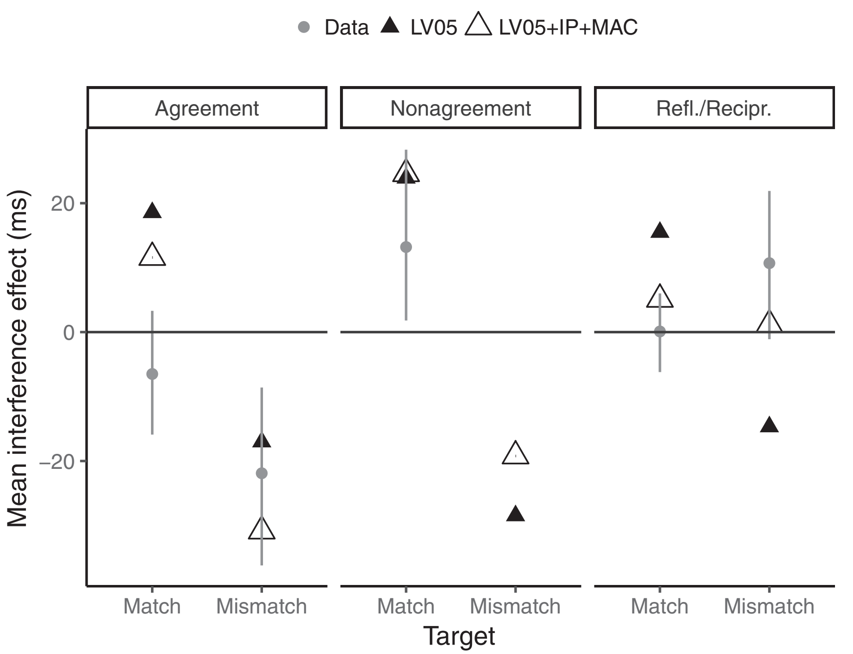



Fig. 10. Mean interference effects from simulations with LV05 and LV05 + IP + MAC for target-match and target-mismatch configurations of the meta-analysis, grouped by dependency type (studies are listed in Table B1 in the Appendix). Human data are shown as mean effect estimates with Bayesian 95% credible intervals as reported in Jäger et al. (2017).

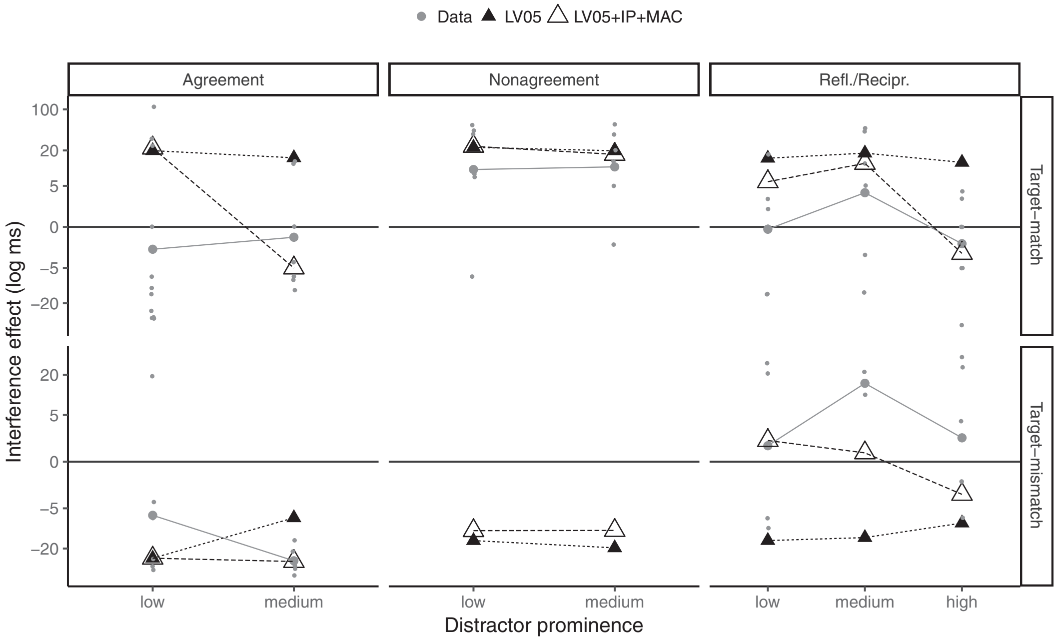



Fig. 11. Mean interference effects from simulations with LV05 and LV05 + IP + MAC for target-match (top panel) and target-mismatch configurations (bottom panel) in the Jäger et al. (2017) meta-analysis, grouped by distractor prominence level within dependency types (studies are listed in Table B1 in the Appendix). Human data are shown as raw means with additional smaller points representing individual studies. The target-mismatch plot in non-agreement subject–verb dependencies does not contain human data because no data were available at the time of the meta-analysis. However, Cunnings and Sturt (2018) have recently found evidence consistent with the predictions of the model; in two experiments, they obtained an estimated mean of −22 ms with a 95% credible interval of [−4, −42], and in a second experiment, a mean of −19 ms, [−40, 1].

the *individual* studies' data, which are classified by dependency type and prominence category (see Table B1). Such a display of the studies' sample means is very different from the estimates from the meta-analysis of Jäger et al. (2017), which summarize what we have learned from the collection of studies on each dependency type. The reason that we are not using the estimates in this figure is that, due to the sparsity of the data, these were not available by prominence level and dependency type in the meta-analysis.

Although the extended model does not show a facilitatory effect in *subject–verb agreement* target-match configurations on average (collapsing over all prominence levels in Fig. 10), Fig. 11 shows that it produces the correct result for those studies that are categorized as having higher distractor prominence. This result of facilitatory interference arises because the prominence parameter $p_{dstr}$ in the LV05 + IP + MAC model was estimated to be higher on average for medium prominence experiments compared to low prominence experiments, as summarized in Table 4.

For non-*agreement subject–verb dependencies*, the fit did not improve in the LV05 + IP + MAC model, because the data only contain target-match configurations, for which the results—mainly inhibitory interference—are perfectly compatible with LV05.

Table 4
Estimated values for prominence parameter pdstr in the LV05 + IP + MAC model with decay for three prominence levels

| Dependency | Low | Medium | High |
|---|---|---|---|
| Agreement | 0.00 | 1.70 | |
| Non-agreement | −0.20 | −0.30 | |
| Reflexives | −1.40 | −1.00 | 4.00 |
| Reciprocals | −1.90 | 0.70 | |

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.

The most interesting results are observed in *reflexive and reciprocal dependencies*. Looking at the means separately for each prominence category shows that the extended model offers an explanation for why the target-match and target-mismatch effects on average seem to deviate from the predictions of LV05. As can be seen in Fig. 11, the average effects in reflexive/reciprocal 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 Fig. 4 shown earlier). Consequently, the extended model matches this pattern while LV05 does not. LV05 + IP + MAC produces a mixture of inhibitory and facilitatory effects in target-match as a consequence of distractor prominence, which would explain why no effect

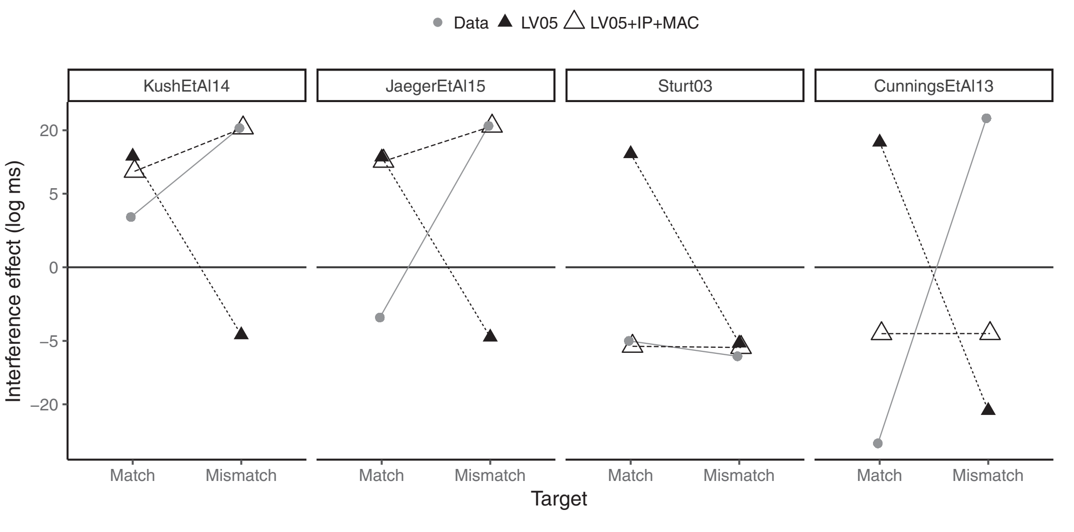



Fig. 12. Human data and simulation results of LV05 and LV05 + IP + MAC for interference effects in target-match and target-mismatch configurations of four individual studies: Kush and Phillips (2014); Jäger et al. (2015, Exp. 1); Sturt (2003, Exp. 1); and Cunnings and Felser (2013, Exp. 2, participants with low working memory).

could be found in the data on average. In target-mismatch configurations, the data show inhibitory effects on average in all three prominence categories. This is incompatible with LV05. LV05 + IP + MAC produces means with a positive sign for low and medium distractor prominence of reflexives and reciprocals. These are driven by the model fitting the inhibitory target-mismatch effects of Kush and Phillips (2014) and Jäger et al. (2015) with an increased estimate of the cross-association parameter for both studies. Thus, the explanation for the inhibitory effect seen on average in reflexive/reciprocal target-mismatch effects in the data would be that some of the studies on reflexive/reciprocal dependencies contained in the meta-analysis qualify for high cue-feature cross-association levels.

Finally, for a better understanding of what drives the differences between the two models, we show four exemplary cases in Fig. 12, where the data *qualitatively* deviate 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 discrimination as explained in the section on multi-associative cues. As a result of the cue-feature cross-association, LV05 + IP + MAC shows *inhibitory* interference effects in target-mismatch configurations, whereas LV05 shows facilitation. The model parameter for the cross-association level was estimated at 0.7 for both reciprocals (Kush & Phillips, 2014) and *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 either of the two tested models. The inhibitory target-mismatch effect is not fit by LV05 + IP + MAC 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 simulate 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 not compatible with any model. We return to this point in Section 5.

## 5. General discussion

The aims of this work were to investigate the quantitative predictions of the Lewis and Vasishth (2005) model and to investigate the consequences of memory accessibility and context-dependent cue-feature associations in the light of the available evidence from reading studies on interference effects in dependency resolution. We have presented an implemented model of prominence and multi-associative cues as an extension to the cue-based retrieval model of LV05. The extension consisted of three revisions of previously simplifying assumptions of ACT-R/LV05 modeling:

1′. The base-level activation of items in memory (i.e., accessibility) is affected by—in addition to recency—their prominence in the current context, that is, their relevance/

salience in terms of syntactic relations in a sentence or information structural and discourse properties.

2′. The strength of any interference effect, that is, also the fan effect, is not simply determined by the presence versus the absence of a distractor but also changes as a function of the distractor's activation in memory relative to the target.

3′. The associative strength between a retrieval cue and a memory item can be the result of multiple cues being associated with multiple features at variable degrees. Cue-feature associations are based on associative learning through language experience.

Our simulations show that prominence and multi-associative cues can account for a range of data points that were not predicted by the original model. In particular, while the prediction space of LV05 allows only two qualitatively different outcomes (inhibition in target-match and facilitation in target-mismatch configurations), the prediction space of the extended model allows, under certain specific circumstances, all four qualitative outcomes seen in the data (inhibition and facilitation in both target-match and target-mismatch configurations). This shows that well-motivated assumptions are of crucial importance when specifying a model, as slight alterations can have consequences not only quantitatively but also qualitatively. In the current case, accounting for individual study design and integrating independently motivated assumptions about memory accessibility and context-based feature discrimination considerably changed the model's prediction space. 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.

It is important to note that the model does not predict just any possible outcome; if that were the case, the model would not be very meaningful or useful (Roberts & Pashler, 2000). First of all, the predictions of prominence and cross-associated cues are restricted to very specific circumstances regarding the grammatical and discourse role of the distractor in the individual experiment and the type of dependency used (e.g., reflexive or reciprocal). The second constraint is that, while some parameters were estimated for best fit with the data in the simulations, parameters were fixed across all conditions of an individual experiment. This restricts the predictions of the model considerably; for example, the model cannot predict, for the same 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 in Exp. 2 of Cunnings and Felser (2013) as shown in Fig. 12. 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. An important line of future work would be to attempt to replicate the Cunnings and Felser result; the model predicts that it should not replicate. 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. Thus, our simulation methodology considerably restricts the model's prediction space and are based on independently motivated assumptions.

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 B1 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 in particular experimental designs. 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.

Some caution is also needed as regards the interpretation of the available data. As discussed in Jäger et al. (2017) and Vasishth et al. (2018), low power and publication bias could be important factors that weaken the empirical claims. Appendix B in 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) and many others before them 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 will be exaggerated in magnitude (Type M error) and can have the wrong sign (Type S error). It would, therefore, be worthwhile to reevaluate the predictions of this extended LV05 model with larger sample studies. For example, how do LV05's predictions fare in target-mismatch reflexives/reciprocals? In English reflexives, if we assume that gender marking on the reflexive *himself/herself* is used as a retrieval cue to seek out an antecedent, the LV05 model predicts facilitatory interference effects in target-mismatch configurations. Dillon et al. (2013) argued that the parser was immune to facilitatory interference based on a 40-subject study. A Bayesian reanalysis of their data Jäger et al. (2019) shows a mean estimate of $-18$ ms, and Bayesian 95% credible interval $[-72, 36]$. This was a fairly low-powered study; as discussed in Appendix A of Jäger et al. (2019), if the true effect size were to be $-23$ ms (the median effect predicted by LV05), then prospective power for a replication of their study would be about 13%. This means that there is an approximately 87% chance of obtaining a non-significant result even though the null hypothesis is false with this particular value for the effect size. Jäger et al. (2019) conducted a larger sample replication attempt (181 subjects); power for the same effect size of $-23$ ms is about 42%. Jäger and colleagues' larger sample study found a facilitatory interference effect of $-23$ ms, 95% credible interval $[-48, 2]$. This estimate turns out to be consistent with the LV05 model's predictions (under the assumption that gender is used as a retrieval cue in

English reflexives). This example illustrates the need for obtaining more precise estimates of the effects of interest than we currently have. In Vasishth et al. (2018), we provide further discussions of this general point about the adverse consequences of low power on developing an empirical base for theory testing, and provide constructive suggestions on how the situation could be improved.

The data on non-agreement subject–verb dependencies agree overall with the general LV05 predictions—inhibition in target-match configurations—and thus had a good fit in 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 are thus not included in our simulations. However, a recent study by Cunnings and 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 eye tracking-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 9a compared to 9b.

(9) 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.

Our own Bayesian estimate of their effect size in their Experiment 1 is $-22$ ms with a credible interval of $[-4, -42]$; for their Experiment 2, the estimate is $-19$ ms $[-40, 1]$. These are consistent with both the original and extended LV05 model's predictions.

To summarize, Table 1 suggested that the LV05 model makes the incorrect predictions for target-mismatch in reflexives and reciprocals, but the Jäger et al. (2019) replication attempt indicates that the LV05 predictions may be correct. Furthermore, the Cunnings and Sturt (2018) data are consistent with the LV05 predictions for target-mismatch configurations in non-agreement subject–verb dependencies.

A major contribution of the present work is that it spells out, for the first time, the predictions of the LV05 model with reference to all the evidence that was available from reading studies at the time of writing. The modeling presented here is highly constrained: (a) 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; (b) the model predictions are restricted by interactions between variables such as prominence, recency, and cue-feature cross-association; and (c) 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's predictions. 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 (a) the available data are sparse with respect to the levels of distractor prominence studied within dependency types and different levels of feature discrimination, (b) there may be confounds between prominence and other factors, and (c) there may be different cognitive processes involved in certain dependency types that the model does not account for. In the following sections, we further discuss the implications of our approach for distractor prominence and cue-feature associations and potential alternatives.

### 5.1. Distractor prominence

In the model we have presented, the prominence of a distractor is a function of its syntactic position and discourse status. 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, for example, of being a grammatical subject. The predictions of syntactic weighting are consistent with our prominence account and are also compatible with ACT-R in general and LV05 in particular. Because of its reduced activation, a distractor that mismatches the subject cue 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. Eq. 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 the activation of both 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, that is, 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 can be extended to discourse status or other contributing factors that we have not considered here: For example, thematic role (Arnold, 2001), contrastive focus (Cowles et al., 2007), first mention (Cowles et al., 2007), and animacy (Fukumura & van Gompel, 2011) are known to affect discourse saliency and might thus influence distractor prominence. Importantly, our account predicts a facilitatory effect in target-match configurations as a consequence of high distractor prominence. This cannot be explained in terms of cue combinatorics.

### 5.2. *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. One way of describing the learning process could be 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 and Wagner (1972) 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 10.

(10) 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 (Fig. 11) 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 one as strong 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 morpho-syntactic 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 (Logacev & 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 fewer 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 has a very large standard error ($M$ = 22 ms, $SE$ = 26 ms), but the sign of the estimated mean is consistent with the assumption of an individual-level variation of cue-feature associations due to adaptive processes. Note, however, that, even if it were 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 and 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 high-powered planned experiments should be carried out in order to test this hypothesis.

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) has proposed a hierarchy between features, such that *person* > *number* > *gender*. Additionally, in English, *number* has a regular, general affixal realization on nouns and verbs, whereas *animacy* and *gender* do not. The effects of semantically, pragmatically, or morphologically motivated differences between retrieval cues remain to be investigated.

### 5.3. Some limitations of the present work

The principal goals of this work are to (a) evaluate the predictions of the Lewis and Vasishth's 2005 (LV05) model against all available reading data, and (b) to propose a plausible account for the datasets that the LV05 model cannot explain. In doing so, we proposed two new constructs, prominence and cue association. Introducing these new constructs obviously raises further questions as to the generality of their application. For example, in our discussion of prominence we have only considered how interference effects play out as a function of prominence, which is, perhaps over-simplistically, limited to subject-hood and discourse topic-hood. We have left underspecified how prominence might work more generally for co-reference resolution. As Kaiser and Trueswell (2008) showed, in a language like Finnish, pronouns and demonstratives exhibit different amounts of sensitivity to word order and syntactic role in determining the antecedent. Specifically, the Finnish pronoun *hän*, "he/she," prefers to choose syntactic subjects as an antecedent regardless of word order, but the demonstrative *tämä*, "this," is sensitive to both word order and syntactic role, so that object–verb–subject order would lead to an approximately equal preference for the object and the subject as antecedent, but subject–verb–object order would lead to a subject preference. Clearly, being able to account for pronouns versus demonstratives in a cue-based architecture requires making the assumption that non-canonical word order makes fronted object nouns more prominent and that subjects are prominent even when they are not in sentence-initial position. Our work in this paper has nothing to say about what happens with non-canonical word order, where information structure plays a crucial role. However, we do not claim to provide a comprehensive theory or model of prominence in this paper.

Similarly, the idea of cue-association is proposed in the context of Hindi reciprocals and Chinese reflexives. How generally applicable is cue-association? Ideally, one should present independent evidence for this proposal, using an experiment design such as Example 10 above. As we have discussed earlier, our proposal should be seen as a tentative one that needs empirical verification through appropriately powered studies.

## 6. 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 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. The assumptions of a continuously valued *fan* as a function of activation as well as a more generalized association between cues and features are independently motivated and domain-general. Their effect could, therefore, be explored in ACT-R modeling also in domains other than sentence processing. 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; Parker, 2019; Rasmussen & Schuler, 2018; Smith et al., 2018) relative to published data.

The web application developed for this paper can be accessed at https://engelmann.shinyapps.io/inter-act/. Researchers are invited to use the application to conduct further simulations.

## Acknowledgments

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, Dorothea Pregla, Sandra Hanne, and Dario Paape. We also thank the audiences of AMLaP 2014, CUNY 2015, and ICCM 2015 for commenting on our work. Our deep gratitude belongs to Whitney Tabor, who has invested a substantial amount of time and effort into the manuscript throughout the review process. 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 B3 (PIs: Ralf Engbert and Shravan Vasishth).

## Notes

1. The dependent measures used in the meta-analysis were first-pass reading times in eye tracking and reading time in self-paced reading at the critical or post-critical region. One reason for using first-pass reading time in the meta-analysis was that in the earliest work on English reflexives (Sturt, 2003), a distinction was made between early vs. late processes, and it was argued by Sturt that in the early stages of processing, indexed by first-pass reading time, reflexives are immune to interference. Thus, the relevant measure for evaluating interference effects in reflexives was first-pass reading times; for consistency, we used first-pass reading times from all other studies on interference as well. Subsequent work related to interference (e.g., Cunnings & Sturt, 2018; Dillon et al., 2013) based their conclusions only on effects found in total reading times. As a consequence, in our more recent evaluations of the LV05 model (e.g., Jäger et al., 2019), we use total reading times.
2. 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.
3. 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. Note that although *local-clause* is also a relational feature, it may still be a useful heuristic to approximate *c-command* since much less computations are needed to keep track of it. See Kush (2013) for an investigation of the computational complexity needed for keeping track of c-commanders.
4. 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 reanalysis process that inflates reading times (McElree, 1993). For an implementation and quantitative

comparison of the direct-access model (McElree et al., 2003) with the LV05 model, see Nicenboim and Vasishth (2018).

5. We acknowledge that the representation of the structural binding requirement as a single cue is a simplification. Theoretically, anaphor binding would require an item to be c-commanding and within the anaphor's binding domain. In some of the studies simulated here, the distractor mismatches both of the requirements and in some studies it mismatches only one of them. However, the number of overloaded cues that stay unchanged across conditions (i.e., match the same items in all conditions) does not affect the predictions because an interference effect arises in the model due to the difference in matched cues between conditions. In the case where the distractor mismatches two structural cues instead of one, the distractor would receive less spreading activation in all conditions. As a consequence, the predicted sizes of the effects would be smaller. Qualitatively, however, the results would not change.
6. Note that, in Fig. 3, there are 334 out of 10,980 simulated data points in target-mismatch configurations that show *inhibitory* interference. These are associated with a specific parameter configuration, namely, with a high retrieval threshold (0) and a low maximum associative strength (1). These outcomes are therefore most likely related to retrieval failures. For this reason, and because the effects are small and make up only 3% of target-mismatch data points, we do not consider inhibitory target-mismatch effects a systematic prediction of LV05.
7. 95% credible intervals are computed within the Bayesian data analysis framework (Gelman et al., 2014). The range specified by a 95% credible interval contains the true value of the estimated parameter with 95% certainty, given the model and the data.
8. The reader may wonder why a facilitatory race effect does not emerge in target-mismatch when cross-associated cues induce a fan effect: after all, both the target and distractor again have similar activations. A similar activation of target and distractor is a prerequisite for a race-based statistical facilitation. For a race-induced facilitation, however, the target would have to be similarly activated in both conditions c and d, as is the case without cross-association (Fig. 1). With cross-association, the target has higher activation in c, whereas in d, the activation is shared with the distractor and therefore reduced. Therefore, the target's activation in c is much higher relative to the activations in condition d of the two racing items (i.e., the target and the distractor). Hence, the target in c will be retrieved faster on average than the winner of the race in condition d, and consequently no facilitation effect will be observed in d vs. c.
9. As discussed in Kush and Phillips (2014), Hindi reciprocals have properties identical to English reciprocals: the antecedent must c-command the reciprocal and also match the reciprocal in morphological features (plural), and the antecedent must be in the same clause as the reciprocal.
10. We thank Klaus Oberauer for his helpful comments, which led to the present implementation.

11. The actual level that leads to detectable inhibitory interference depends on the specific situation being simulated and the values of other ACT-R parameters.
12. 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."

## 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), 1036–1060.

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

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

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). Studies in Generative Grammar. 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. *Journal of Memory and Language*, *102*, 16–27.

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.

Dotlacil, J. (2018). Building an ACT-R reader for eye-tracking corpus data. *Topics in Cognitive Science*, *10* (1), 144–160.

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). Studies in Discourse and Grammar. Amsterdam, The Netherlands: John Benjamins.

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

Engelmann, F., & Vasishth, S. (2019). Inhibitory and facilitatory interference. figshare. Figure. https://doi.org/10.6084/m9.figshare.9305456.v2

Engelmann, F., Vasishth, S., Engbert, R., & Kliegl, R. (2013). A framework for modelling 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., & vanGompel, 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.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Chapman and Hall/CRC.

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 modelling 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.

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2019). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, *23*(5), 709–748.

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

Kush, D. (2013). Respecting relations: Memory access and antecedent retrieval in incremental sentence processing. 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*. Stanford, CA: Stanford University Press.

Lewis, R. L., & Badecker, W. (2010).Short-term memory in sentence production and its adaptive control.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. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

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

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*, *10*(1), 161–174.

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., Logacev, 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).

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34.

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*, *42*, 1075–1100.

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*(1), 15–30.

Parker, D. (2019). Cue combinatorics in memory retrieval for anaphora. *Cognitive Science*, *43*(3), e12715.

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.

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).

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*. Vienna, Austria: R Foundation for Statistical Computing.

Raab, D. H. (1962). Division of psychology: Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*(5 Series II), 574–590.

Rasmussen, N. E., & Schuler, W. (2018). Left-corner parsing with distributed associative memory produces surprisal and locality effects. *Cognitive Science*, *42*, 1009–1042.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Reinhart, T. M. (1976). The syntactic domain of anaphora.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: Appleton-Century-Crofts.

Roberts, S., & Pashler, H. (2000). 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. *Cognitive Science*, *42*, 1043–1074.

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: Harvard University Press.

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. A. (2002). Parsing as working memory retrieval: Interference, decay, and priming effects in long distance attachment.Doctoral dissertation, University of Pittsburgh, PA.

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

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

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

Vasishth, S., Bruessow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712.

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175.

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*. Submitted.

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

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.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1: Supplementary Material.

## Appendix A: Key terms and concepts

Table A1
Terminology used in the present article in relation to cue-based retrieval and interference in dependency resolution

| Term | Definition |
|---|---|
| Feature | Any property of an item represented in memory<br>Example: the representation of the lexical item *girl* has features *animate* and *female* |
| Retrieval cue | A feature used to seek out an item in memory for retrieval<br>Example: the retrieval cue *animate* is used to seek out the subject of *laughed* |
| Relevant cues | The retrieval cues that are part of the experimental manipulation |
| Target | The item that is the correct target for retrieval |
| Distractor | An item that is not the correct target for retrieval |
| Match | A match occurs when a retrieval cue and a feature on an item have the same value |
| Mismatch | A mismatch occurs when a retrieval cue and a feature on an item do not have the same value |
| Full match | All relevant retrieval cues (usually two) are matched by the features of an item |
| Partial match | Some but not all (usually one of two) retrieval cues are matched by the features of an item |
| Target-match | Target-match configurations are sentences where the target matches all relevant retrieval cues |
| Target-mismatch | Target-mismatch configurations are sentences where the target does not fully match the relevant retrieval cue(s) |
| Interference | The effect of a (partially) matching distractor on the retrieval of the target |
| Interference condition | (*distractor-match*) Manipulation of a target-match or target-mismatch sentence such that a distractor matches at least one of the retrieval cues |
| No-interference condition | (*distractor-mismatch*) Manipulation of a target-match or target-mismatch sentence such that no distractor matches any relevant retrieval cues |
| Inhibitory effect | A slowdown in processing during retrieval of the target due to interference from a distractor |
| Facilitatory effect | A speed-up in processing during retrieval of the target due to interference from a distractor |
| Activation | The strength with which an item is represented in memory. More highly activated items are easier to access, resulting in more accurate and/or faster retrieval (depending on the theory) |
| Base-level activation | A function of an item's time of creation, its intermittent reactivations and time-based decay |
| Spreading activation | The activation boost that a memory item receives as the result of a match with one or more retrieval cues |
| Cue overload | This occurs when a retrieval cue matches the features of two or more items. The cue is ambiguous |
| Misretrieval | The retrieval of a distractor rather than the target |
| Fan effect | Reduction in activation of items in memory as a result of other items matching the same retrieval cue |
| Statistical facilitation | A speed-up in average processing time caused by random noise in a race between two similarly activated items |
| Interference effect | |

(continued)

*Table A1 (continued)*

| Term | Definition |
|---|---|
| | The difference in processing time (retrieval latency) between the interference and the no-interference condition (distractor-match − distractor-mismatch). The effect is positive (i.e., slow-down or inhibition) when processing in the interference condition is slower than in the no-interfere condition, and negative (i.e., speed-up or facilitation) when processing is faster |

Table A2
Terminology used in the present article in relation to our extension of the cue-based retrieval model

| Term | Definition |
|---|---|
| Prominence | Elevated activation of an item in memory, caused by factors unrelated to the retrieval cues, for example, grammatical position or discourse marking |
| Cue-feature association | Assuming that the feature value of an item does not have to be identical with the retrieval cue in order to produce a match, the cue-feature association level determines how strong the match between a retrieval cue and a feature is |
| Feature co-occurrence | Two features are called co-occurring in a certain retrieval context when the combination of both features identifies the correct target more often than other feature combinations |
| Cross-association | As the result of feature co-occurrence, two retrieval cues can become cross-associated in the sense that both cues are associated with—and therefore produce a match with—the same features to a certain degree |
| Feature discrimination | A retrieval cue is highly discriminative if it is associated with only one (or very few) features. A retrieval cue is less discriminative if it is associated with multiple features. Low feature discrimination is the result of *feature co-occurrence* and can lead to *cross-association* |

## Appendix B: List of experiments included in the simulations

Table B1
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 |

(continued)

*Table B1 (continued)*

| Dependency | Prominence | ID | Publication | Int. Type | Lang. | Distr. Pos. |
|---|---|---|---|---|---|---|
| | | 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-agreement | Low | 19 | Van Dyke and McElree (2006) | Pro | EN | 3× memory |
| | | 20 | Van Dyke and McElree (2011, E2b) | Pro | EN | Obj |
| | | 21 | Van Dyke (2007, E1, LoSyn) | Retro | EN | PP |
| | | 22 | Van Dyke (2007, E3, LoSyn) | Retro | EN | PP |
| | | 23 | Van Dyke (2007, E2, LoSyn) | Retro | EN | PP |
| | | 24 | Van Dyke and McElree (2011, E2b) | Retro | EN | Obj |
| | Medium | 25 | Van Dyke and McElree (2011, E1bpro) | Pro | EN | Subj |
| | | 26 | Van Dyke and McElree (2011, E1bretro) | Pro | EN | Subj |
| | | 27 | Van Dyke (2007, E1, LoSem) | Retro | EN | PP, subj |
| | | 28 | Van Dyke (2007, E2, LoSem) | Retro | EN | PP, subj |
| | | 29 | Van Dyke (2007, E3, LoSem) | Retro | EN | PP, subj |
| | | 30 | Van Dyke and Lewis (2003, E4) | Retro | EN | PP, subj |
| Reciprocals | Low | 31 | Kush and Phillips (2014) | Retro | HI | Prepobj |
| | Medium | 32 | Badecker and Straub (2002, E4) | Pro | EN | Subj |
| Reflexives | Low | 33 | Badecker and Straub (2002, E5) | Pro | EN | Gen |
| | | 34 | Badecker and Straub (2002, E6) | Pro | EN | Prepobj |
| | | 35 | Jäger et al. (2015, E2) | Pro | CN | 3× 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 and Straub (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 and Felser (2013, E1, HiWMC) | Pro | EN | Subj, topic |
| | | 45 | Cunnings and Felser (2013, E1, LoWMC) | Pro | EN | Subj, topic |
| | | 46 | Cunnings and Sturt (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 |

(continued)

*Table B1 (continued)*

| Dependency | Prominence | ID | Publication | Int. Type | Lang. | Distr. Pos. |
|---|---|---|---|---|---|---|
| | | 49 | Sturt (2003, E1) | Pro | EN | Subj, topic |
| | | 50 | Cunnings and Felser (2013, E2, HiWMC) | Retro | EN | Subj, topic |
| | | 51 | Cunnings and Felser (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).

## Appendix C: Model specifications

Table C1
Model parameters, their default values, and the values used in the simulation of the studies in the meta-analysis

| Parameter | Name | Default | Simulation |
|---|---|---|---|
| $F$ | Latency factor | 0.2 | [0.1, 0.25] |
| $f$ | Latency exponent | 1 | 1 |
| $\tau$ | Retrieval threshold | −1.5 | −1.5 |
| $d$ | Decay constant | 0.5 | 0.5 |
| $ANS$ | Activation noise | 0.2 | 0.2 |
| $MAS$ | Maximum associative strength | 1 | 1.5 |
| $MP$ | Mismatch penalty | 1 | 0.25 |
| $\beta$ | Base-level constant | 0 | 0 |
| $t_{trgt}$ | Time since last target presentation | 1,000 | {700, 1,300} |
| $t_{dstr}$ | Time since last distractor presentation | 1,000 | {700, 1,300} |
| Extended parameters | | | |
| $q$ | Match quality correction factor | 10 | 0, 10 |
| $c$ | Cross-association level | 0 | [0, 1] |
| $p_{trgt}$ | Target prominence | 0 | 0 |
| $p_{dstr}$ | Distractor prominence | 0 | [−2.5, 4] |

Table C1 lists all model parameters with their default values and the values used in the simulation of the studies in the meta-analysis.

Eqs. C1–C5 specify model components that were not defined in the text or were represented in a simplified way.

The noise component (Eq. C1) that is part of the activation function in Eq. 4 is a normally distributed random variable scaled by the noise parameter *ANS*. The mismatch penalty component (Eq. C2), which is usually part of the base-level activation function (here in its complete form in Eq. C3) assigns a penalty for every cue $j$ that item $i$ does not match. The complete equation for the retrieval time of item $i$ (Eq. C4) is a function of that item's activation $A_i$ if $A_i$ is equal to or above the retrieval threshold $\tau$, and is a function of $\tau$ otherwise. Noise is then added (Eq. C5) by transforming the retrieval time $RT$ into a uniformly distributed random variable $\widehat{RT}$ within the range of $\pm\frac{1}{3}RT$.

$$\epsilon_i \sim \text{Normal}\left(\mu = 0,\, \sigma = \sqrt{\frac{\pi^2}{3}ANS^2}\right) \quad \text{noise} \tag{C1}$$

$$Penalty_i = MP \times \sum_{j=1}^{n} (P(i|j) - 1) \quad \text{mismatch penalty} \tag{C2}$$

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) + \beta_i + Penalty_i + p_i \quad \text{base} - \text{level} \tag{C3}$$

$$RT_i = \begin{cases} Fe^{-fA_i}, & \text{if } A_i \geq \tau \\ Fe^{-f\tau} & \text{otherwise} \end{cases} \quad \text{retrieval time} \tag{C4}$$

$$\widehat{RT}_i = \text{Uniform}\left(\frac{2}{3}RT_i,\, \frac{4}{3}RT_i\right) \quad \text{noisy retrieval time} \tag{C5}$$