William & Mary
CHARTERED 1693

W&M ScholarWorks

Arts & Sciences Articles

Arts and Sciences

1-30-2019

Cue combinatorics in memory retrieval for anaphora

Dan Parker

Follow this and additional works at: https://scholarworks.wm.edu/aspubs

Part of the Psycholinguistics and Neurolinguistics Commons

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 43 (2019) e12715
© 2019 Cognitive Science Society, Inc. All rights reserved.
ISSN: 1551-6709 online
DOI: 10.1111/cogs.12715

Cue Combinatorics in Memory Retrieval for Anaphora

Dan Parker

Department of English, Linguistics Program, College of William & Mary

Received 17 April 2018; received in revised form 23 January 2019; accepted 30 January 2019

Abstract

Many studies have shown that memory retrieval for real-time language processing relies on a cue-based access mechanism, which allows the cues available at the retrieval site to directly access the target representation in memory. An open question is how different types of cues are combined at retrieval to create a single retrieval probe ("cue combinatorics"). This study addresses this question by testing whether retrieval for antecedent-reflexive dependencies combines cues in a linear (i.e., additive) or nonlinear (i.e., multiplicative) fashion. Results from computational simulations and a reading time experiment show that target items that match all the cues of the reflexive are favored more than target items that mismatch these cues, and that different degrees of mismatches slow reading times in comparable amounts. This profile is consistent with the predictions of a nonlinear cue combination and provides evidence against models in which all cues combine in a linear fashion. A follow-up set of simulations shows that a nonlinear rule also captures previous demonstrations of interference from nontarget items during retrieval for reflexive licensing. Taken together, these results shed new light on how different types of cues combine at the retrieval site and reveal how the method of cue combination impacts the accessibility of linguistic information in memory.

Keywords: Memory retrieval; Sentence processing; Anaphora; Cue combinatorics; Reading times; Computational modeling

1. Introduction

It has long been known that working memory shapes the dynamics of real-time language processing (e.g., Chomsky & Miller, 1963; Fodor, Bever, & Garrett, 1974; Frazier & Fodor, 1978; Jarvella, 1971; Jarvella & Herman, 1972; Kimball, 1973, 1975; McElree, 2000; McElree, Foraker, & Dyer, 2003; Miller & Chomsky, 1963; Miller & Isard, 1963; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006,

Correspondence should be sent to Dan Parker, Department of English, College of William & Mary, P.O. Box 8795, Williamsburg, VA 23187-8795. E-mail: dparker@wm.edu

2007, 2011; Wundt, 1904). The role of working memory in language processing is evident in contexts involving the so-called long-distance dependencies, which require a comprehender to relate two nonadjacent constituents for interpretation. Common examples of long-distance dependencies include subject-verb number agreement, as in (1a), and reflexive-antecedent dependencies, as in (1b).

(1) a. **The children** at the park **were** playing on the slide.
b. **The girl** at the park hurt **herself** on the slide.

Successful interpretation of the sentences in (1) requires matching a specific form (e.g., the agreeing verb *were* or reflexive anaphor *herself*) to a specific item (a "licensor" or "antecedent") that appears earlier in the sentence (e.g., *the children* or *the girl*). How this matching process happens has been the subject of much debate in psycholinguistics. One possibility is that comprehenders actively maintain a representation of the target item across the span of the dependency. But given the stringent limits on the amount of material that can concurrently occupy working memory, active maintenance of the target is not always feasible (Cowan, 2001; Jonides et al., 2008; McElree, 2006; McElree & Dosher, 1989; Miller, 1956). To remedy these limitations, comprehenders may engage memory retrieval mechanisms to recover the necessary information for interpretation.

Evidence from a range of methodologies, constructions, and languages suggests that memory retrieval for dependency formation relies on a *content-addressable* memory access mechanism (Dillon, Chow, & Xiang, 2016; Dillon, Mishler, Sloggett, & Phillips, 2013; Foraker & McElree, 2007; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Martin & McElree, 2008, 2009, 2011; Martin, Nieuwland, & Carreiras, 2012; McElree, 2000; McElree et al., 2003; Parker & Lantz, 2017; Parker & Phillips, 2016, 2017; Tanner, Nicol, & Brehm, 2014; Tucker, Idrissi, & Almeida, 2015; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006, 2011; Vasishth, Brüssow, Lewis, & Drenhaus, 2008; Wagers, Lau, & Phillips, 2009). On this view, syntactic dependencies like those in (1) are formed using a retrieval process in which all of the cues available at the retrieval site are assembled into a single retrieval probe that is matched against all task-relevant items in memory based on their content features (Clark & Gronlund, 1996; Kohonen, 1980). The item that provides the best match is retrieved to establish the dependency (modulo effects of stochastic noise, cue overload, and similarity-based interference, which will be discussed later).

For instance, encountering the reflexive anaphor *herself* in (1b) triggers a retrieval process in which a set of syntactic and semantic cues corresponding to the syntactic position, person, number, gender, and animacy features of the antecedent (e.g., +subject, +3rd person, +singular, +feminine, +animate) are combined into a single retrieval probe that is matched against the corresponding content features of each item in the sentence (Parker & Phillips, 2017). In (1b), the item *the girl* provides the best match to the cues of the retrieval probe and therefore should, in principle, be retrieved as the antecedent.

This account of memory retrieval for sentence comprehension is attractive because it is derived from independently motivated principles of working memory attested across

perceptual and cognitive domains (Lewis, 1996; McElree, 2000; McElree et al., 2003; see Caplan & Waters, 2013, and Jonides et al., 2008, for a review). In particular, it has been claimed that the same set of working memory principles that governs general cognition is applied in retrieval for sentence comprehension (Lewis & Vasishth, 2005; Lewis et al., 2006; McElree, 2000; McElree et al., 2003), motivating a unified theory of memory retrieval processes across cognitive domains.

Many of the cue-based retrieval models described in the cognitive science literature assume that the probability of retrieving an item in memory is a function of the strength of association between each cue in the retrieval probe and the corresponding content features of the memory item (for a review, see Caplan & Waters, 2013; Clark & Gronlund, 1996, or Jonides et al., 2008). Specifically, each cue has a strength of association with each item in memory, taking into account "cue overload" or the degree to which the cue matches other items in memory (Nairne, 2002). The overall strength of association between the target and retrieval probe (i.e., probe-to-target match score) reflects the combined strengths of association for all of the cues in the retrieval probe. On this view, items that match more cues will have a higher probability of retrieval and will be integrated back into the current processing stream more quickly than items that match fewer cues. Thus, probe-to-target match score is a key determinant of the success and timing of retrieval.

Research in this domain has identified two ways in which the individual cue strengths can be combined at retrieval. One method of combining cue strengths is with a *linear* (i.e., additive) function, such that each cue contributes directly to an item's overall match score, independently of the strengths of the other cues in the retrieval probe. This combination method predicts that for a given item, its match score (and hence, probability of retrieval) will grow linearly as the number of matching cues increases (e.g., Trommershäuser, Körding, & Landy, 2011, pp. 7–8). Cue strengths can also combine in a *non-linear* (i.e., multiplicative) fashion, such that the contribution of each cue is not independent, but rather depends on the overall degree of match with the other cues in the retrieval probe. With this method, the total contribution of the matching cues will exceed their simple summation, resulting in a superadditive/exponential growth in an item's match score (and retrieval probability) as the number of matching cues increases (e.g., Trommershäuser et al., 2011, pp. 10–17).[1]

There is a large psychophysics literature examining the application of these cue combination methods within a wide range of sensory modalities (see Trommershäuser et al., 2011, for an introduction). However, cue combinatorics has received far less attention in the domain of psycholinguistics and has only recently begun to be studied systematically. For instance, Martin (2016) presents a novel model of language processing in which cue combination and integration serves as the link between psycholinguistic theory and neurobiological models of language, grounding psycholinguistic process models in canonical neurophysiological computation. Martin's model focuses on the broad issue of cue integration across representational levels (e.g., phonemes, syllables, morphemes, words, phrases, syntactic structures, and discourse context), where each level of representation acts as a cue to higher levels of representation, resulting in a "cascaded" architecture for

language processing. In this model, it is assumed that parsing and other language processing phenomena can be accounted for using two important psychophysiological operations: *summation*, the main mechanism for cue combination; and *normalization*, which incorporates probabilistic estimates of a cue's reliability (Martin, 2016, p. 10). Martin's model is the first of its kind to forefront the role of cue combination and integration in the transition between levels of representation for language processing.

More narrowly in terms of retrieval for dependency formation during real-time sentence processing, the prominent ACT-R model of sentence processing developed by Lewis and Vasishth (2005) assumes that retrieval mechanisms always apply a linear cue combination method uniformly for all dependencies (e.g., eq. 2 in Lewis & Vasishth, 2005). In this model, the notion of a cue is much more restricted than that which is assumed in Martin's (2016) cue-integration model. At the sentence level, cues are features derived from the current word, linguistic context, and grammatical knowledge, and form a subset of the features of the target item, that is, the antecedent/licensor (Lewis et al., 2006, p. 448), are combined to retrieve the target as needed. In the ACT-R implementation of cue-based retrieval, it is simply stipulated that cues are combined in a linear fashion.

Previous research in the cognitive and perceptual domains has shown that both a linear and nonlinear cue combination method are needed to explain perceptual behavior, even within the same cognitive domain. In other words, the two methods are not mutually exclusive (see Trommershäuser et al., 2011, for a review). For instance, psycholinguistic studies have presented data that are consistent with a linear cue combination method in retrieval for dependency formation, as well as data that are consistent with a nonlinear cue combination method (e.g., Van Dyke & McElree, 2011; see also Parker & Phillips, 2017). However, models of memory retrieval in sentence processing have typically adopted either a linear or nonlinear method exclusively for all cases of retrieval, and it remains unclear under what conditions a linear vs. nonlinear cue combination method is applied in retrieval for dependency formation. As Van Dyke and McElree (2011) explain, a complete account of memory retrieval in sentence processing must explicitly characterize the types of cues that guide retrieval and how those cues combine at retrieval for language processing. This study contributes to filling this gap using both computational and behavioral methods, focusing on cue combinatorics in retrieval for anaphor resolution as a model test case.

1.1. Cue combination methods

The notion of cue-based retrieval is often made explicit within an "activation-based" memory architecture. In an activation-based architecture, items encoded in memory are differentially activated based on their probe-to-target match score, such that items with a higher match score have a higher activation value, resulting in a higher probability and a faster processing latency. Memory models that adopt a linear cue combination method define activation according to Eq. 1 (e.g., Anderson, 1990; Anderson et al., 2004; Lewis & Vasishth, 2005; Vasishth et al., 2008). Eq. 1 states that the activation A_i for an item I_i

is the summation of strength of association S between each retrieval cue Q_j and the features of the item, expressed as $W_jS(Q_1, I_i)$, where W_j reflects the weight associated with the cue. Most implementations in psycholinguistics assume that cues are weighted equally (see Kush, 2013, for discussion).

$$A_i = \sum_{j=1}^{n} W_j S(Q_j, I_i) \tag{1}$$

The most important feature of Eq. 1 for the present purpose is that the strength of association for each cue contributes directly to the item's activation, independently of the strengths of association for the other cues in the retrieval probe. Since the function is additive, activation grows linearly with each matching cue (for an illustration, see Fig. 1 and Fig. 2 in the next section). This feature permits activation of items that match some, but not all, of the cues in the retrieval probe ("partial matches"), increasing the probability that nontarget items will interfere with retrieval of the target.

By contrast, memory models that adopt a nonlinear cue combination rule define activation according to Eq. 2 (e.g., Gillund & Shiffrin, 1984; Hintzman, 1984; Nairne, 1990; Raajimakers & Shiffrin, 1981; Van Dyke, 2007; Van Dyke & McElree, 2006). Unlike with a linear rule, the contribution of individual cues in a nonlinear combination is not independent. Retrieval exhibits sensitivity to conjunctions of cues, rather than the occurrence of individual cues, such that target items that match all of the cues are favored more than partially matching target items, with all degrees of partial matches being

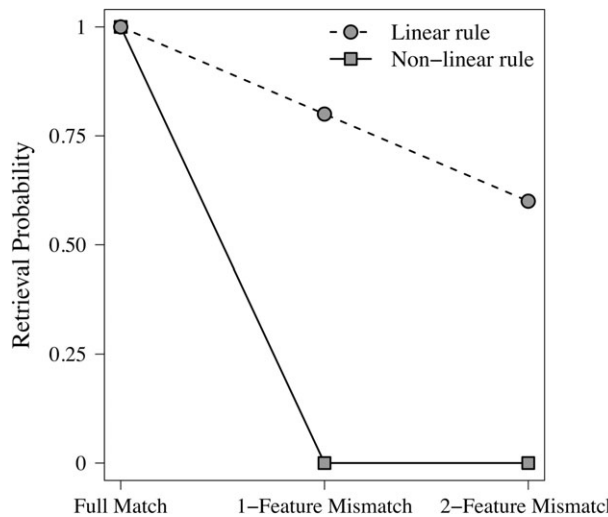

Fig. 1. Predicted retrieval probabilities for the target item by condition for the linear and nonlinear rules from Experiment 1.

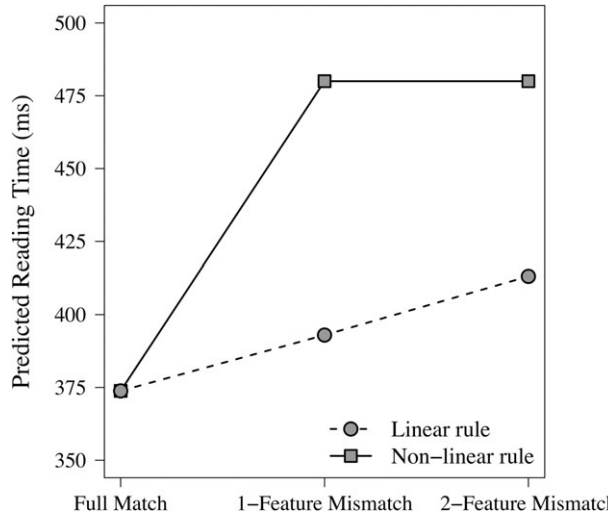

Fig. 2. Predicted reading times for the linear and nonlinear cue combination rules from Experiment 1.

disfavored in comparable amounts. This is because cue strengths are multiplied, rather than summed, which causes a much greater reduction in activation for partial matches than occurs with a linear scheme (see Fig. 1 and Fig. 2 in the next section for an illustration). This feature makes interference from nontarget partial matches less likely, relative to the linear combinatorics scheme (see Fig. 6 in Experiment 3).

$$A_i = \prod_{j=1}^{n} S(Q_j, I_i)^{w_j} \tag{2}$$

Trommershäuser et al. (2011) explain that while most studies on cue combination in the cognitive and perceptual domains have found evidence consistent with a linear model (e.g., see chapters 1–3 for a review), there are certain perceptual phenomena that motivate a nonlinear model (e.g., see chapters 4–5 for a review), and there are even cases where linear and nonlinear methods are jointly needed to explain behavior within the same domain, modality, and perceptual circumstance (e.g., see pp. 33–34 for a brief summary).

In the domain of sentence processing, evidence of interference from nontarget partial matches during retrieval for linguistic dependency formation is consistent with the predictions of a linear combinatorics scheme. For instance, many studies on subject-verb agreement processing have shown that nontarget partial matches can disrupt retrieval of the target subject at the verb (e.g., example 1a; Dillon et al., 2013; Lago et al., 2015; Tanner et al., 2014; Tucker & Almeida, 2017; Tucker et al., 2015; Wagers et al., 2009). Such effects are frequently observed in ungrammatical sentences where a plural verb matches a plural noun that is not the target subject in configurations like (2).

(2) *The key to **the cabinets** apparently **were** on the table.

According to retrieval-based accounts of agreement processing (e.g., Dillon et al., 2013; Lago et al., 2015; Tanner et al., 2014; Tucker & Almeida, 2017; Tucker et al., 2015; Wagers et al., 2009), encountering the plural verb *were* in (2) triggers a retrieval process that seeks an item matching the cues +*subject* and +*plural*. In (2), the target noun phrase (NP) *the key* matches the +*subject* cue, but it does not match the +*plural* cue. By contrast, the NP *the cabinets* does not match the +*subject* cue, but it does match the +*plural* cue. At retrieval, the partial feature match on +*plural* boosts the activation of the memory representation of *the cabinets*, making it likely to interfere with retrieval of the target if its activation passes a certain threshold due to stochastic noise. For present purposes, the finding that individual features like +*plural* influence retrieval, independently of the match to the other cues, is consistent with the predictions of a linear cue combination.

However, not all cases of interference clearly implicate a linear combination. For instance, Van Dyke and McElree (2011) examined the effects of retrieval for subject-verb thematic binding in configurations like (3a,b). In both (3a) and (3b), the verb *compromised* requires an animate subject, motivating the use of +*animate* and +*subject* as retrieval cues. In (3a), the NP *the witness* matches both of these cues, but in (3b), it only matches the +*animate* cue, because it is located in a direct object position.

(3) a. The attorney who the judge realized had declared that **the witness/**the motion was inappropriate compromised during the negotiations.
b. The attorney who the judge realized had rejected **the witness/**the motion in the case compromised during the negotiations.

Van Dyke and McElree (2011) found that the presence of the animate subject in (3a) disrupted reading times at the verb, but the animate object in (3b) did not, in comparison to their respective baseline conditions with an inanimate distractor. Van Dyke and McElree (2011) argued that these results are consistent with a linear combination, but one in which syntactic cues are given greater weighting than semantic cues, such that syntactically inappropriate object distractors in sentences like (3b) have little or no measurable impact on retrieval. However, they are explicit that their results are also consistent with a nonlinear combination, in which only items that match all the cues, that is, both +*animate* and +*subject*, are considered at retrieval. Similarly, Parker and Phillips (2017) recently reported findings of selective interference effects for antecedent–reflexive dependencies, which are consistent with either a weighted linear rule or a nonlinear rule (see also Cunnings & Sturt, 2014).

The conditions under which a linear or nonlinear method is applied in retrieval for sentence comprehension remain unclear. In particular, existing data from interference studies make it difficult to distinguish when linear and nonlinear cue combinations apply, since previous research on retrieval in sentence processing has focused on a narrow set of cue combinations. In particular, most studies on retrieval for linguistic dependency formation have been limited to tests of a single feature manipulation involving either gender, number, or

animacy (see Parker & Phillips, 2017, for discussion), making it difficult to assess how different types of cues are assembled to create the retrieval probe. Addressing the question of cue combinatorics in sentence processing is important to better understand the factors that cause memory retrieval to succeed or fail during language comprehension.

1.2. The present study

This study uses a combination of behavioral and computational methods to tease apart the predictions of the linear and nonlinear cue combination methods. The most important difference between linear and nonlinear combination methods for present purposes is their sensitivity to target items that fully match the cues relative to target items that match only a subset of the cues (i.e., the conjunction of cues). This difference entails divergent predictions for dependency formation during sentence comprehension. For instance, in the case of reflexive-antecedent dependencies like (1b), a nonlinear combination rule predicts that a target item that matches all the cues should be favored exponentially more than a target item that matches only a subset of the cues. A linear combination rule, by contrast, predicts an additive effect that increases linearly with each additional matching cue.

This study tests these predictions to determine which cue combination scheme is used in retrieval for antecedent-reflexive dependencies—a model test case for examining the computational properties of memory retrieval (Chen, Jäger, & Vasishth, 2012; Cunnings & Felser, 2013; Cunnings & Sturt, 2014; Dillon, 2014; Dillon, Chow, & Xiang, 2016; Jäger, Engelmann, & Vasishth, 2017; Kush & Phillips, 2014; Parker & Phillips, 2017; Patil, Vasishth, & Lewis, 2016; Sturt, 2003; Xiang, Dillon, & Phillips, 2009). Previous studies on retrieval for reflexive licensing have not addressed the issue of cue combinatorics, as they focused on the broad question of whether partial matching nontarget items interfere with retrieval of the target. However, given that the presence or absence of interference effects is not always a clear indicator of the underlying cue combination method (see Van Dyke & McElree, 2011), the question of what cue combination method is applied for antecedent-reflexive dependencies remains open. To address this issue, this study begins by focusing on contexts without a distractor in Experiments 1 and 2, which manipulated the degree of antecedent-reflexive match to test the predictions regarding cue convergence for the linear and nonlinear combination methods. Lastly, the question of how the candidate combination methods play out for contexts with a distractor is taken up in Experiment 3.

The paper is organized in the following manner. Experiment 1 used computational modeling to generate precise quantitative predictions about the timing of dependency formation for the linear and nonlinear combination rules, using reflexive-antecedent dependencies like those in (1b) as a model test case. Experiment 2 tested the model's predictions by manipulating the degree of match between the retrieval cues of the reflexive and the antecedent, comparing target items that either fully matched the retrieval cues (full match), had a single feature mismatch (1-feature mismatch), or had two feature mismatches (2-feature mismatch) using reading time measures. Results showed that target items that matched all the cues (full match) were favored more than target items with a 1- and 2-feature mismatch, and crucially, that the 1- and 2-feature mismatches slowed

reading times in comparable amounts. This profile is consistent with a nonlinear cue combination and provides evidence against models that assume that cues always combine in a linear fashion, such as the ACT-R model, validating the need for future research on cue combinatorics. Experiment 3 provides a follow-up set of simulations that outline the predictions of the candidate cue combination methods for interference paradigms involving a feature-matching distractor and shows that a nonlinear combination method can capture previous demonstrations of interference in antecedent retrieval for reflexive licensing (e.g., Parker & Phillips, 2017). These results provide further evidence that a nonlinear rule applies in retrieval for reflexive licensing. Taken together, the results of Experiments 1–3 suggest that the method of cue combination is a key determinant of target accessibility during retrieval for sentence comprehension.

2. Experiment 1: Computational modeling

Experiment 1 used computational modeling to generate precise quantitative reading time predictions for the linear and nonlinear cue combination functions in Eqs. 1 and 2. The model used antecedent-reflexive dependencies like those in (1b) to assess how each of the candidate cue combinations would play out at retrieval. Antecedent-reflexive dependencies provide a model case for testing assumptions about cue combinatorics for several reasons. First, many studies have shown that encountering a reflexive anaphor like *herself, himself,* or *themselves* during real-time comprehension triggers a memory retrieval for an antecedent in the previous context (Badecker & Straub, 2002; Clifton, Frazier, & Deevy, 1999; Cunnings & Felser, 2013; Cunnings & Sturt, 2014; Dillon et al., 2013; Nicol & Swinney, 1989; Parker & Phillips, 2017; Runner, Sussman, & Tanenhaus, 2006; Sturt, 2003; Xiang et al., 2009). Second, reflexive-antecedent dependencies have strict linguistic constraints that define a precise target for retrieval. For instance, the antecedent must agree in person, number, and gender with the reflexive, and in the constructions tested in this study, the antecedent must be the subject of the clause that contains the reflexive (Chomsky, 1981; Reinhart, 1976; Reinhart & Reuland, 1993). These constraints can be viewed as instructions for the retrieval mechanism to find an item with specific features in a specific location, and recent work has verified that these constraints actively guide antecedent retrieval, with a direct mapping between the overt linguistic features and retrieval cues (Parker & Phillips, 2017). Third, and most important for present purposes, reflexive-antecedent dependencies allow independent manipulation of the match between the retrieval cues at the reflexive and the corresponding features of the antecedent. For instance, it is possible to systematically manipulate the degree to which the antecedent matches these cue (i.e., cue convergence), as shown in Table 1. This property is necessary to test the competing cue combination methods, as it allows investigation of how each of the cue combination methods behaves with respect to varying degrees of partial matching targets.

One concern about this design is that some of the conditions involve a violation due to the mismatch between the reflexive and target, which means that the results for these conditions might not reflect "normal" processing. While there are limitations to the

Table 1
Sample items for Experiments 1 and 2

Reflexive-Antecedent Feature (Mis)match	Sentence
Full match	The man recently hurt himself at the park
1-feature mismatch (gender)	*The woman recently hurt himself at the park
2-feature mismatch (gender and number)	*The women recently hurt himself at the park

Note. The symbol "*" indicates that the sentence is ungrammatical due to a reflexive-antecedent feature mismatch.

interpretation of data from violation paradigms, the current design is necessary to tease apart the candidate cue combination methods, which are distinguished by their predictions regarding the effects of cue convergence. The current design also facilitates comparison with previous studies. The majority of the existing studies on retrieval in sentence comprehension have relied on a violation paradigm to draw inferences about the underlying principles of the retrieval mechanisms (see Jäger et al., 2017, for a recent review), and these studies have shown that manipulating the match between the cues and the items in memory can reveal how cue-match guides retrieval. This study follows this tradition, using the effects of cue convergence to gain insights about cue combinatorics.

2.1. Method

2.1.1. Procedure

The linear and nonlinear combination rules shown in Eqs. 1 and 2 were implemented in the R software environment (R Development Core Team, 2018) for retrieval in sentences with a reflexive-antecedent dependency using cues corresponding to the target subject position, person, number, and gender (following Wagers, 2008). Predictions for the sensitivity to (mis)matching targets were generated for the candidate cue combination rules by systematically manipulating the degree to which the cues converged on the target (full match, 1-feature mismatch, 2-feature mismatch), as shown in Table 1.

Each matching (i.e., convergent) cue was assigned a high strength of association (1.00), and each mismatching cue was assigned a low strength of association (0.00), following Wagers (2008). From these values, the resulting activation value for the target item from each condition was normalized to a full match (i.e., normalized to unity), in which all cues converged, to calculate the retrieval probabilities for each target item. Most cue-based memory models assume that the probability of retrieving an item is proportional to its activation, that is, A_i in Eqs. 1 and 2 (see Clark & Gronlund, 1996; Lewis & Vasishth, 2005), and normalizing to unity is a standard procedure for converting continuous values, such as activation values, into probabilities [0, 1] (Grus, 2015).

The activation values for each condition were then mapped to reading time predictions according to Eq. 3, which is a widely adopted function for mapping activation onto reading times (Bothell, 2007; Lewis & Vasishth, 2005). T reflects the time to recover an item i and integrate it back into the processing stream in milliseconds. F and f are two scaling constants

that set the predictions on an appropriate time scale to fit the dependent measures (e.g., self-paced reading times). Broadly, Eq. 3 states that as activation decreases, T increases.

$$T_i = Fe_i^{-(f \times A_i)} \tag{3}$$

To facilitate comparison with the reading time measures that will be obtained in Experiment 2, it is necessary to spell out some linking assumptions regarding the relationship between the latencies generated by Eq. 3 and reading time measures. This study adopts the standard linking assumption that the latencies generated by the model are monotonically related to the reading time measures that index retrieval operations, such that longer latencies entail longer RTs (Anderson & Milson, 1989). There are certainly additional contributors to the observed RTs, such as interpretation time, and sometimes reanalysis. But this study assumes with many others (e.g., Anderson, Budiu, & Reder, 2001; Boston, Hale, Vasishth, & Kliegl, 2011; Dillon, Chow, Wagers, et al., 2016; Dillon et al., 2013; Jäger, Engelmann, et al., 2015; Kush & Phillips, 2014; Lewis & Vasishth, 2005; Nicenboim, Logačev, Gattei, & Vasishth, 2016; Nicenboim & Vasishth, 2018; Patil et al., 2016; Tucker et al., 2015; Vasishth et al., 2008) that these additional processes do not disrupt the monotonic relation between retrieval and reading times. Importantly, these studies have shown that the reading times predicted by Eq. 3 provide a good quantitative fit to the reading time data obtained in behavioral studies on retrieval in sentence comprehension.

2.1.2. Results and discussion

Fig. 1 shows the predicted retrieval probabilities as a function of target cue convergence for the linear and nonlinear combination rules. For the linear rule, a full matching target has a relatively high probability of retrieval, which decreases linearly as the number of cue mismatches increases. The nonlinear rule, by contrast, predicts sensitivity to the conjunction of cues, but not the occurrence of individual cues, such that a full matching target is favored more than mismatching targets, with both the 1-feature and 2-feature mismatches being disfavored in comparably.

Fig. 2 shows how the competing retrieval functions map to reading time predictions according to Eq. 3. Both combination methods predict increased reading times for the mismatching targets relative to the full match. However, it is the shape of the reading time disruption that distinguishes the two methods: the linear rule predicts an additive effect, such that each additional cue mismatch yields a linear increase in reading times, whereas the nonlinear rule predicts a sharp increase in reading times for the mismatching targets relative to the full match, such that the different degrees of mismatches increase reading times in comparable amounts.

3. Experiment 2: Testing the model's predictions

The goal of Experiment 2 was to test the model predictions for the linear and nonlinear cue combination rules using self-paced reading for reflexive-antecedent dependencies

like those in Table 1. Importantly, many studies have shown that self-paced reading measures reliably index retrieval operations during real-time dependency formation (e.g., Chen et al., 2012; Dillon, Chow, & Xiang, 2016; Jäger, Benz, Roeser, Dillon, & Vasishth, 2015; Kush & Phillips, 2014; Paape, 2016; Parker, Lago, & Phillips, 2015; Slioussar & Malko, 2016; Tucker & Almeida, 2017; Tucker et al., 2015; Van Dyke & McElree, 2006; Wagers et al., 2009; Xiang, Grove, & Giannakidou, 2013). To this end, Experiment 2 fit the processing disruptions generated by the (mis)matching targets to the model predictions generated in Experiment 1 to determine which cue combination is used to guide retrieval for reflexive processing.

3.1. Method

3.1.1. Participants

One hundred twenty-six native speakers of American English were recruited using Amazon's Mechanical Turk Web service (https://www.mturk.com). The number of participants was determined by a statistical power analysis using the data generated by the computational model in Experiment 1. The model predicted a 20 ms difference between the two mismatch conditions for the linear model. Assuming a standard deviation of 75 ms (following Jäger, Engelmann, et al., 2015; Jäger, Benz, et al., 2015, who assumed identical values for self-paced reading measures), the power analysis suggested that at least 150 participants would be needed to achieve 90% probability of detecting this effect. Funding restrictions limited the actual sample size to 126 participants. As validation, the observed standard deviation was slightly smaller at 63 ms, which yielded a statistical power of 94% with 126 participants. All participants provided informed consent and were screened for native speaker abilities. The screening probed knowledge of the constraints on English tense, modality, morphology, ellipsis, and syntactic islands. Participants in Experiment 2 were compensated $3.00. The experiment lasted approximately 20 min.

3.1.2. Materials and design

The experimental materials consisted of 18 item sets of the form shown in Table 1. Across all conditions, the subject noun phrase served as the antecedent for the reflexive, which always appeared as the direct object of the verb. The reflexive was followed by a 3–6 word spillover region. The reflexive always appeared in the third-person, singular form (e.g., *himself* or *herself*). The form of the reflexive remained constant across conditions within each item set. The antecedent systematically varied in the degree of match relative to the features of the reflexive (i.e., cue convergence), along the dimensions of number and gender (full match, 1-feature mismatch, 2-feature mismatch).[2] Item sets were balanced such that nine sets contained masculine reflexives (*himself*) and nine sets contained feminine reflexives (*herself*). Following previous studies (Dillon et al., 2013; Parker & Phillips, 2017; Sturt, 2003), the target antecedent used both stereotypical gender (e.g., *nurse*) and definitional gender (e.g., *mother*).

The 18 item sets were divided into three lists in a Latin square design and combined with 36 grammatical filler sentences of similar length and complexity compared to the

test items, yielding an overall grammatical-to-ungrammatical ratio of 3.5:1 (previous studies on reflexives range from 1.8:1 to 4.6:1; Xiang et al., 2009 and Dillon et al., 2013, respectively). To balance the number of grammatical-to-ungrammatical reflexives, 1/3 of the fillers contained grammatical reflexives. The full list of experimental materials can be found in Data S1. All items were followed by a comprehension question that addressed various parts of the sentence to prevent the possibility that participants would read only the material needed to answer the question.

3.1.3. Procedure

The experiment was conducted using the online experiment platform Ibex (http://spellout.net/ibexfarm), which allows self-paced reading experiments to be deployed in a standard web browser. Sentences were initially masked by dashes, with white spaces and punctuation intact. Participants pushed the space bar to reveal each word. Presentation was noncumulative, such that the previous word was replaced with a row of dashes when the next word appeared. Each sentence was followed by a "yes/no" comprehension question, and an onscreen notification was provided for incorrect answers. The order of presentation was randomized for each participant.

To ensure that participants completed the task as directed, an instructional manipulation check was used (Oppenheimer, Meyvis, & Davidenko, 2009). Instructional manipulation checks ensure that participants are completing the task as directed by asking them to confirm that they read the instructions. For this study, the instructional manipulation check required participants to respond to a set of comprehension questions following the instructions (e.g., *What button should I press to advance a word?* and *What button do I press to respond "Yes" to a comprehension question?*). Two participants were excluded from the analysis due to failure to respond correctly to the instructional manipulation check, yielding a total of 124 participants for analysis.

3.1.4. Analysis

Average reading times were compared across conditions in the following regions of interest: the word immediately before the reflexive ("pre-critical"), the reflexive ("critical"), and the following word ("spillover"). Statistical analyses were carried out over the raw reading times as well as the log-transformed values to control for the log-normal distribution of reading times (Box & Cox, 1964; Ratcliff, 1993; Vasishth & Nicenboim, 2016). Models were constructed using the *lme4* package (Bates, Maechler, & Bolker, 2011) in the R software environment (R Development Core Team, 2018). Contrast coding was applied to examine the effects of cue match (C1: full match vs. 1-feature mismatch) and cue mismatch (C2: 1-feature mismatch vs. 2-feature mismatch) at each region of interest. Both comparisons were included in the same model as fixed effect predictors. All models were fitted with a full variance–covariance matrix (i.e., a maximal random effects structure), with random intercepts and slopes for participants and items (Baayen, Davidson, & Bates, 2008; Bates et al., 2011). If the model failed to converge or the variance–covariance matrix was degenerate (e.g., correlations were close to ± 1), random slopes for participants and then items were removed until convergence obtained

(i.e., the "BS Backward" procedure described in Barr, Levy, Scheepers, & Tily, 2013). The structure of the final (i.e., converging) models is as follows:

Precritical region: lmer(RT ~ C1 + C2 + (1 + C1|Item) + (1|Subject), data = df)
Critical region: lmer(RT ~ C1 + C2 + (1 + C2|Item) + (1 + C1 + C2|Subject), data = df)
Spillover region: lmer(RT ~ C1 + C2 + (1 + C2|Item) + (1 + C1 + C2|Subject), data = df)

A fixed effect was considered significant if its absolute t value was >2, which indicates that its 95% confidence interval did not include 0 (Gelman & Hill, 2007). Reading time data were then compared to the model predictions for the linear and nonlinear cue combination rules generated in Experiment 1 using the adjusted R^2 statistic.

3.1.5. Predictions

According to the model predictions generated in Experiment 1, both cue combination methods predict a significant effect of cue match, but only the linear method predicts an effect of cue mismatch, with significantly longer reading times for the 2-feature mismatch condition relative to the 1-feature mismatch condition. The nonlinear method, by contrast, predicts no difference between the 1-feature and 2-feature mismatch conditions.

3.1.6. Results

Fig. 3 shows the mean reading time data by region and condition. Fig. 4 shows the effect of cue convergence at the critical region, and Fig. 5 shows the same for the spillover region. Figures with the logged reading times are provided in Data S1. No effects were observed in the precritical region (all ts < 0.15). At the reflexive region, there was a main effect of cue match, carried by increased reading times for the 1-feature mismatch condition relative to the full match condition ($\hat{\beta} = 0.06$, $SE = 0.01$, $t = -3.53$). No differences were observed between the mismatch conditions with respect to cue mismatch ($\hat{\beta} = -0.00$, $SE = 0.01$, $t = -0.04$). The same profile was observed at the spillover region (cue match: $\hat{\beta} = 0.16$, $SE = 0.02$, $t = -6.84$; cue mismatch: $\hat{\beta} = -0.00$, $SE = 0.02$, $t = -0.04$). Statistical analyses over the raw reading times showed the same patterns of significance. A comparison with the model predictions generated in Experiment 1 revealed that the nonlinear cue combination rule provided a better fit to the observed reading time data in both the reflexive region (adjusted R^2 for the linear rule = 0.49; adjusted R^2 for the nonlinear rule = 0.99) and spill-over region (adjusted R^2 for the linear rule = 0.48; adjusted R^2 for the nonlinear rule = 0.99). These measures were based on the values shown in Fig. 4 and Fig. 5, respectively.

3.1.7. Discussion

Experiment 2 tested the effect of cue convergence in retrieval for reflexive licensing and compared the observed profiles with the model predictions generated in Experiment 1 to determine whether retrieval for reflexive licensing uses a linear or nonlinear cue combination rule. Results revealed two findings. First, reading times were modulated by

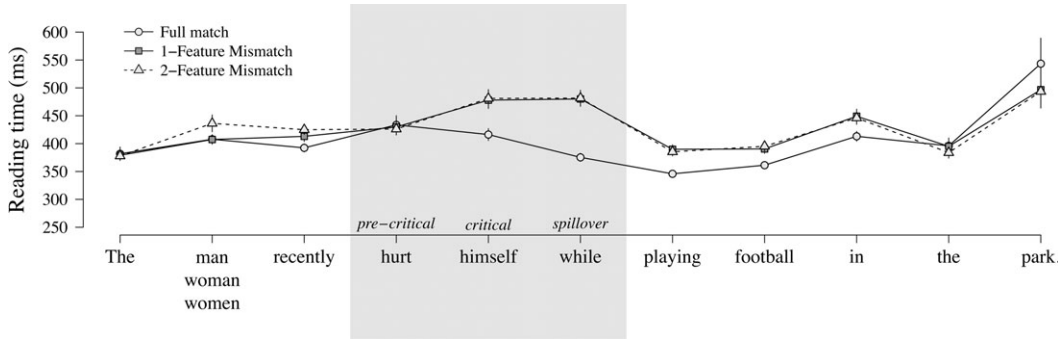

Fig. 3. Mean reading times by region and condition for Experiment 2. Error bars indicate standard error of the mean.

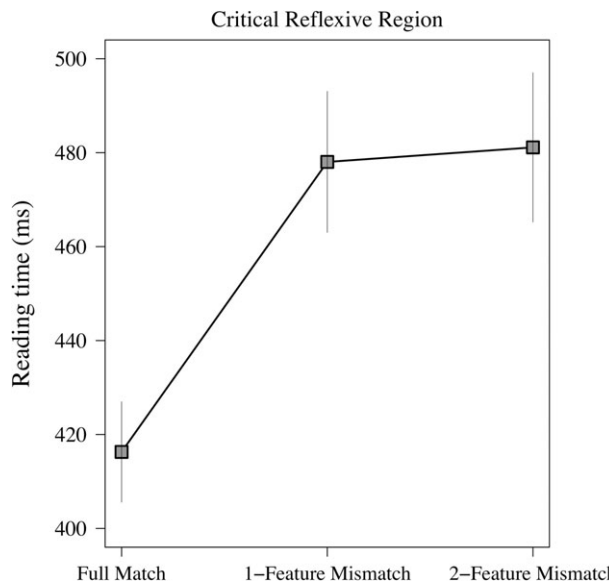

Fig. 4. The effect of cue convergence at the critical region for Experiment 2.

feature mismatches, such that sentences with a target item that matched all the cues of the reflexive were read more quickly at the reflexive than sentences with a target item that mismatched the cues. This effect replicates the basic effect of feature match reported in previous studies on reflexive processing (e.g., Dillon et al., 2013; Parker & Phillips, 2017; Sturt, 2003) and demonstrates that comprehenders were sensitive to the feature (mis)matches used in the present study. Second, and most important for present purposes, the 1- and 2-feature mismatches slowed reading times in comparable amounts at the reflexive and spillover regions. These results are consistent with the predictions of a non-linear cue combination from the computational modeling experiment (Experiment 1). By

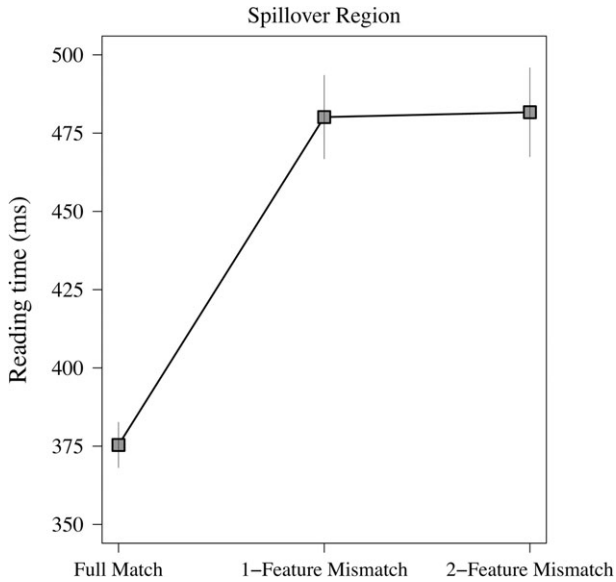

Fig. 5. The effect of cue convergence at the spillover region for Experiment 2.

contrast, these results are not compatible with models that assume that cues always combine linearly, such as the ACT-R model of sentence processing (Lewis & Vasishth, 2005).

A concern with the results of Experiment 2 is that the processing disruption for the mismatching targets might reflect a ceiling effect. For instance, reading times at the reflexive region might have shown a nonlinear trajectory not because retrieval deployed a nonlinear cue combination rule, but rather because the processing disruptions associated with cue mismatches had peaked, which might have masked potential differences between mismatch conditions. There are several reasons why the nonlinear profile at the reflexive is not a ceiling effect. First, the spillover region shows that the processing disruption associated with the cue mismatches can reach even more extreme values than those observed at the critical reflexive region. For instance, at the critical reflexive region, the 2-feature mismatch resulted in a 64 ms disruption, relative to the Full match, whereas at the spillover region, the 2-feature mismatch resulted in a 106 ms disruption. This profile suggests that the reading time disruption at the critical reflexive region was not at ceiling.

Second, results of a post hoc distributional analysis indicate that the observed nonlinear profile is not bounded to the ceiling. If the observed nonlinear profile reflects a ceiling effect, we would expect the nonlinearity to be restricted to the most extreme values in the right tail (i.e., the ceiling) of the reading time distribution. To test this possibility, the 50th percentile of reading times at the critical and spillover regions were analyzed to get a sense of how the reading times patterned in the leftward portion of the reading time distribution. Crucially, the results of this analysis presented in Table 2 show that the nonlinear profile observed in the grand mean is not restricted to the right tail, as nonlinearity is observed in

Table 2
Grand means and the 50th percentile values of reading times by condition at the critical reflexive region and the spillover region for Experiment 2

	Full Match	1-Feature Mismatch	2-Feature Mismatch
Critical region			
Grand mean	416	478	481
50th percentile	368	376	376
Spillover region			
Grand mean	375	480	481
50th percentile	344	390	408

Reading times are in milliseconds.

the leftward portion of the reading time distribution. These results provide quantitative evidence that the observed nonlinear profile is not a ceiling effect. Taken together, the current results are most consistent with the predictions of a nonlinear cue combination rule.

4. Experiment 3: Extending the model's predictions to an interference paradigm

Experiment 2 showed that target items that matched all the cues ("full match") were favored more than target items with a 1- and 2-feature mismatch, and crucially, that the 1- and 2-feature mismatches slowed reading times in comparable amounts, as predicted by a nonlinear cue combination rule. However, most existing studies on retrieval for reflexive licensing have focused on the conditions under which partially matching nontarget distractors interfere with retrieval of the target. A notable difference between those studies and Experiments 1 and 2 of this study is that the stimuli from Experiments 1 and 2 did not contain distractors. The decision to not use distractors in Experiments 1 and 2 was made based on the current research goal, which was to better understand how cue combinatorics affects target accessibility, as this is where the candidate combination methods made clear predictions. Experiment 3 sought to extend the findings from Experiments 1 and 2 by examining the predictions of the candidate cue combinatorics schemes for configurations with a feature matching distractor in an interference paradigm.

Recent research has shown that interference effects arise during retrieval for reflexive licensing if specific conditions are met. For instance, Parker and Phillips (2017) showed that interference arises when the target antecedent is a particularly poor match to the retrieval cues. They tested sentences like those in Table 3 using eye-tracking and found that reflexives are susceptible to interference from partially matching distractors, but only selectively, such that interference effects arose when the target mismatched the reflexive in two features, e.g., gender and number (2-feature mismatch), but not when the target mismatched the target in just one feature, for example, gender or number (1-feature mismatch), or matched all of the cues (full match).

The design used in Experiment 2 of this study is similar to that used by Parker and Phillips (2017) in that it also manipulated the degree of match between the reflexive and

Table 3
Examples stimuli from Parker and Phillips (2017)

Full match	The librarian$_{\text{distractor}}$ said that the schoolgirl$_{\text{target}}$ reminded **herself** ...
1-feature mismatch	The librarian$_{\text{distractor}}$ said that the schoolboy$_{\text{target}}$ reminded **herself** ...
2-feature mismatch	The librarian$_{\text{distractor}}$ said that the schoolboys$_{\text{target}}$ reminded **herself** ...

the target. However, there is a salient contrast between the current findings and those reported by Parker and Phillips (2017). Parker and Phillips (2017) found that the 2-feature mismatch condition patterns differently from the 1-feature mismatch condition (with respect to interference effects), whereas this study showed that the 1-feature and 2-feature mismatch conditions patterned similarly (with respect to effects of target match). Both of these effects reflect cue-based retrieval, but the source of the 1- vs. 2-feature contrast with respect to interference effects remains unclear. To address this issue, Experiment 3 provides additional simulations to investigate whether a linear or nonlinear cue combination could predict the categorical 1- vs. 2-feature mismatch contrast reported in Parker and Phillips (2017) to better understand the nature of cue combinatorics in retrieval for sentence comprehension.

4.1. Method

4.1.1. Procedure

In a cue-based retrieval architecture, the likelihood of interference is a function of "cue overload," or the degree to which the retrieval cues match the target relative to the match to the other items in memory. This phenomenon can also be described in terms of "cue diagnosticity," or the ability of the cues to uniquely identify the target: As the cues match more items in memory, they become less diagnostic to the target and the likelihood of retrieving the wrong item increases (Martin, 2016; Martin et al., 2012; McElree, 2000, 2006; McElree et al., 2003; Nairne, 2002; Van Dyke & McElree, 2006). More formally, the likelihood of interference can be expressed as the strength of association between the cues and the target divided by the summation of the strengths of association between the cues and the other items in memory (Nairne, 2002; see also Martin et al., 2012). Experiment 3 implemented this formalization for the linear and nonlinear combination methods, as shown in Eqs. 4 and 5, to generate predictions about the likelihood of interference for the sentences in Table 3. The model used the same parameters as in Experiment 1.

$$P(I_i|Q_1\ldots,Q_n) = \frac{\sum_{j=1}^{n} W_j S(Q_j, I_i)}{\sum_{k=1}^{N}\sum_{j=1}^{n} W_j S(Q_j, I_k)} \tag{4}$$

$$P(I_i|Q_1\ldots,Q_n) = \frac{\prod_{j=1}^{n} S(Q_j, I_i)^{w_j}}{\sum_{k=1}^{N}\prod_{j=1}^{n} S(Q_j, I_k)^{w_j}} \tag{5}$$

4.1.2. Results and discussion

Fig. 6 shows the probability of interference (i.e., retrieval of the feature-matching distractor) for the linear and nonlinear cue combination methods as a function of the degree of target match for the configurations shown in Table 3. The nonlinear combination predicts comparably low probabilities of interference for the full match and 1-feature mismatch conditions, with a sharp increase in probability of interference in the 2-feature mismatch condition. This profile aligns well with the findings reported in Parker and Phillips (2017), which showed that interference occurs when the target mismatched the reflexive in two features, but not when it mismatched in one feature or fully matched the reflexive. The linear rule, by contrast, failed to predict this profile, showing similar rates of interference across conditions, with a small linear increase in interference with each mismatch. Overall, the predictions of the nonlinear cue combination rule provided a better fit to the total reading times at the reflexive region reported in Parker and Phillips (adjusted R^2 for the nonlinear rule = .99; adjusted R^2 for the linear rule = .61).

The finding that a nonlinear cue combination method can capture previous findings regarding interference effects in reflexive licensing lends additional support to the current proposal that retrieval for reflexive licensing utilizes a nonlinear cue combination method. However, as discussed by Van Dyke and McElree (2011), it may still be possible to capture the observed profiles with a linear rule with extreme differential weighting on syntactic features. Nevertheless, the current results provide more comprehensive evidence that a linear rule alone (without independent motivation for cue weighting) cannot capture the effects of retrieval for reflexive licensing.

5. General discussion

Cue combinatorics has received much attention in the psychophysics literature (Trommershäuser et al., 2011). However, relatively little is known about how cues combine to access linguistic representations in memory during real-time language processing. Recently, it has been claimed that the activation of information across different levels of representation for language processing can be accounted for using the neurobiological mechanisms for cue combination and cue integration (Martin, 2016). This study advanced Martin's theory of cue combination by investigating how cues combine to retrieve information from within a single level of representation, namely the syntactic level, for anaphor processing.

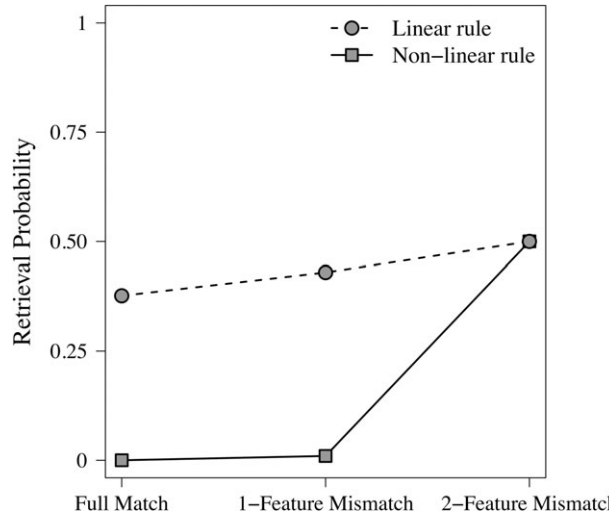

Fig. 6. Probability of interference (i.e., retrieval of the feature-matching distractor) for the linear and nonlinear cue combination methods as a function of the degree of target match, based on the feature matches in Table 3.

5.1. Summary of findings

Research on memory retrieval in the cognitive and perceptual domains has identified two ways in which cues can be combined to access information in working memory, involving a linear and nonlinear cue combination rule. These rules are not mutually exclusive, and both rules can be applied within the same cognitive domain (Trommershäuser et al., 2011). However, existing models of memory retrieval for sentence comprehension have assumed that cues combine either in a linear or nonlinear fashion for all instances of retrieval, and it remains unclear under what conditions the candidate cue combination methods are used in retrieval for language processing.

To address this issue, this study used computational modeling to generate precise quantitative predictions about the timing of dependency formation for the linear and nonlinear combination rules, using reflexive-antecedent dependencies as a model test case (Experiment 1), and tested the model's predictions by manipulating the degree of match between the retrieval cues of the reflexive and the antecedent (Experiment 2). Results showed that target items that matched all the retrieval cues (full match) were favored more than target items with a 1- and 2-feature mismatch, and that the 1- and 2-feature mismatches slowed reading times in comparable amounts. These results are consistent with the predictions of a nonlinear rule, as shown in Fig. 7.

A follow-up set of simulations in Experiment 3 extended the findings from Experiments 1 and 2 by examining the predictions of the linear and nonlinear cue combination methods for configurations with a feature-matching distractor in an interference paradigm. Specifically, Experiment 3 tested whether a linear or nonlinear rule could explain

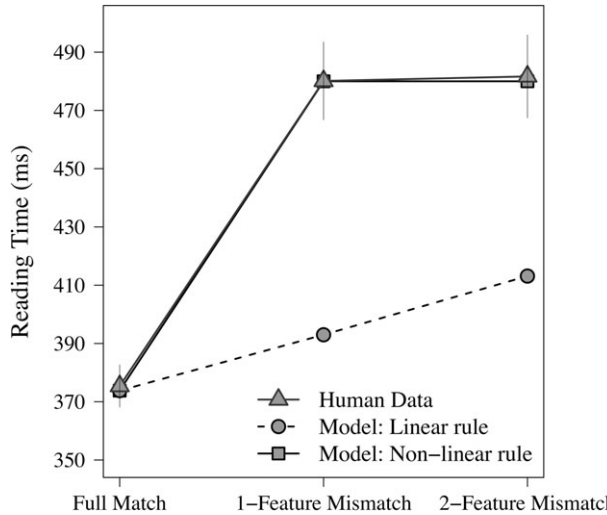

Fig. 7. Comparison of model predictions (Experiment 1) and behavioral data (Experiment 2).

previous findings of selective interference effects in retrieval for reflexive licensing (Parker & Phillips, 2017). Results showed that a nonlinear rule could capture the interference effects reported in Parker and Phillips (2017), and that a linear rule, by contrast, could not. These findings provide further evidence that retrieval for reflexive licensing most likely utilizes a nonlinear cue combination method.

Taken together, the results of Experiments 1–3 suggest that the method of cue combinatorics is a key determinant of target accessibility. Broadly, these results shed new light on how different types of cues combine at retrieval, how the method of combination impacts the accessibility of the target in memory, and how different combination methods determine the success of retrieval in scenarios with a distractor. These findings are consistent with the predictions of Martin's (2016) cue-integration theory, which claims that interference depends on how diagnostic a combined cue set is to the target (described in terms of optimal cue combination). Specifically, the current results demonstrate that the precise method of cue combination that is applied in retrieval for dependency formation affects how cues interact with the contents of memory, reflected in differences of target vs. distractor accessibility as a function of cue diagnosticity (full match vs. 1-feature mismatch vs. 2-feature mismatch).

A concern with the current proposal is that the empirical data could reflect a linear rule with differential cue weighting. Although existing models of retrieval in sentence comprehension, such as ACT-R (Lewis & Vasishth, 2005), have assumed that all cues are combined with equal priority or weighting, Van Dyke and McElree (2011) explain that it is possible for a linear rule to mimic a nonlinear rule if certain cues are implemented with a sufficiently large differential weighting. For instance, it may be possible to achieve the observed nonlinear profile if the gender cue in the 1-feature mismatch condition was strongly weighted relative to the number cue. However, there are several reasons why

such an account cannot be extended to reflexives. First, there is no a priori reason to believe that gender would be more diagnostic or weighted more than number, at least in English. If anything, gender might be less diagnostic than number in the case of reflexive licensing because gender is probabilistically encoded on the antecedent (e.g., stereotypical nouns like *nurse* or *assistant* are more likely to be feminine), whereas number is categorical (e.g., ±singular).

Second, findings across previous studies on reflexive processing suggest that gender and number behave similarly with respect to retrieval for reflexive processing. For instance, Experiment 2 of Dillon et al. (2013) investigated the impact of number on antecedent retrieval for reflexive licensing and found that a number mismatch between the target antecedent and reflexive produced a ~74 ms disruption in total reading times, which is comparable to the ~71 ms disruption associated with a gender mismatch reported in Patil et al. (2016) in a similar structural configuration. This comparison suggests that comprehenders do not appear to be any more or less sensitive to a gender mismatch than they are to a number mismatch. However, verification of this claim for future work depends on a within-participants direct comparison using maximally similar sentences for gender and number mismatches. Third, even if gender was weighted more than number under a linear rule, we would expect the addition of a number mismatch in the 2-feature mismatch condition to further increase the processing disruption. However, no such effect was observed, as reading times for the 1-feature and 2-feature mismatch conditions patterned similarly, as expected with a nonlinear combination method in which cues are weighted equally. Importantly, the current results are not incompatible with the claim that other types of long-distance dependencies weight cues differentially, like the thematic-binding relations tested by Van Dyke and McElree (2011). But some reasons for why certain dependencies might differ with respect to cue combinatorics and cue weighting are considered in the next subsection.

A second concern with the current results is that there are limitations on the insights that we can glean about the retrieval mechanisms based on reading time data. The use of reading time measures in this study was motivated by the fact that prominent past studies relied on reading time data to advance the cue combination theories tested here (e.g., Lewis & Vasishth, 2005; Van Dyke, 2007; Van Dyke & McElree, 2006; Vasishth et al., 2008). However, reading time measures reflect a mixture of cognitive processes, including retrieval, interpretive processes, and sometimes reanalysis, and these additional processes can obscure or mask the relationship between the underlying retrieval procedures and overall reading times. Currently, there does not exist a comprehensive theory of how retrieval and interpretation are jointly reflected in total reading times (in particular, one that is explicit enough to make quantitatively precise predictions). The absence of such a theory motivates the use of linking assumptions. This study adopted the standard linking assumption that the latencies generated by the model are monotonically related to the reading time measures that are considered to index retrieval operations, such that longer latencies entail longer RTs (Anderson & Milson, 1989), and assumed with many others (e.g., Anderson et al., 2001; Boston et al., 2011; Dillon, Chow, Wagers, et al., 2016; Dillon et al., 2013; Jäger, Engelmann, et al., 2015; Kush & Phillips, 2014; Lewis & Vasishth, 2005; Nicenboim

& Vasishth, 2018; Nicenboim et al., 2016; Patil et al., 2016; Tucker et al., 2015; Vasishth et al., 2008) that any additional processes reflected in reading measures do not disrupt the monotonic relation between retrieval and reading times. However, it is possible that the relationship between retrieval and reading times is non-monotonic, and this possibility cannot be excluded without additional tests that do no conflate predictions about differences in processing speed with predictions about differences in representation strength, such as those involving a speed-accuracy trade-off (SAT) procedure (Reed, 1973). Ultimately, examining the ways in which the model's predictions align or diverge from human behavior will put us in a better position to more explicitly characterize the relation between retrieval and interpretation as reflected in reading time measures. This study takes a step in that direction, and the finding that the model provides a good fit to the data (see Fig. 7) suggests that the model and accompanying linking assumptions are on the right track.

5.2. Relation to previous findings on cue-based retrieval in sentence comprehension

The findings from Experiment 3 showed that a nonlinear rule can capture previous demonstrations of interference in retrieval for reflexive licensing. These results raise the question of whether other cases of interference in retrieval for dependency formation can be captured similarly with a nonlinear rule. One challenge for a uniform account is that not all dependencies are equally susceptible to interference. For instance, reflexives generally do not exhibit interference in 1-feature mismatch configurations (Parker & Phillips, 2017; see also Clifton et al., 1999; Dillon et al., 2013; Nicol & Swinney, 1989; Sturt, 2003), which is consistent with the predictions of a *nonlinear* cue combination, but numerous studies have shown that subject-verb agreement dependencies are highly susceptible to interference in contexts with a 1-feature mismatch where the target mismatches the number cue of the verb, for example, **The key to the cabinets are rusty*, which is consistent with a *linear* cue combination (Clifton et al., 1999; Dillon et al., 2013; Lago et al., 2015; Patson & Husband, 2015; Pearlmutter, Garnsey, & Bock, 1999; Staub, 2009, 2010; Tanner et al., 2014; Tucker & Almeida, 2017; Tucker et al., 2015; Wagers et al., 2009).

There are several reasons why subject-verb agreement and reflexive-antecedent dependencies might show different profiles with respect to interference effects. One possibility suggested by Dillon et al. (2013) is that there are dependency-wise differences in cue combinatorics methods. For instance, reflexives might utilize a nonlinear combination by default, but subject-verb agreement, by default, might utilize a linear rule. The possibility that dependencies differ in their cue combination methods is consistent with the findings in the cognitive and perceptual domains showing that linear and nonlinear cue combination methods are not mutually exclusive, and that both are needed to explain perceptual behavior within the same cognitive domain (Trommershäuser et al., 2011).

One relevant factor that motivates the use of a linear rule for subject-verb agreement dependencies is the need for reanalysis. For instance, several researchers have argued that cue-based retrieval for agreement processing functions as a reanalysis mechanism that is engaged when the top-down expectations about the agreement features of the

verb generated by the target subject conflict with the bottom-up information (e.g., Dillon et al., 2013; Lago et al., 2015; Parker & Phillips, 2017). The retrieval mechanisms might selectively engage a linear cue combination in response to this violation to broaden the search for a number-matching item by permitting partial matches, leading to interference on the basis of a single feature mismatch between the target subject and verb. This account lends itself to a sort of noisy-channel analysis in which the parser operates over an uncertain input (Gibson et al., 2013; Halle & Stevens, 1959, 1963; Levy, 2008; Poeppel & Monahan, 2010) and draws on the computational rationality framework (e.g., Lewis, Howes, & Singh, 2014), in which the exact cue combination method engaged for language processing would be determined by the problem at hand. On this view, both a linear and nonlinear cue combination are available for retrieval in sentence comprehension, which is consistent with the claims in the broader psychophysics literature (Trommershäuser et al., 2011), and the need for reanalysis or repair might be one factor that determines which method is applied. An important task for future research is to investigate how different retrieval triggers (e.g., normal processing vs. processing error) impact cue combinatorics and to test a broader range of dependencies to better understand the boundary conditions for the application of the candidate cue combination methods.

Relatedly, future work must also investigate the effect of cue reliability on retrieval for sentence processing. Cue reliability plays a crucial role in Martin's (2016) cue-integration model of sentence processing, which focuses on the integration of information across different levels of representation. In Martin's model, cue integration refers to the weighting of individual cues by estimates of their reliability, based on prior experience and expectations. However, far less is known about how cue reliability impact retrieval on a narrower scale in terms of retrieval for sentence processing. If retrieval for sentence processing is a skilled-based procedure, as previously claimed (e.g., Lewis & Vasishth, 2005), then retrieval could be optimized as a function of language use to deploy only the most frequent and reliable cues to recover the target. Although it is typically assumed that all cues are weighted equally in memory retrieval, some recent studies have argued that certain cues, such as structural cues, might be weighted more heavily than nonstructural cues in retrieval for dependency formation (e.g., Cunnings & Sturt, 2014; Dillon et al., 2013; Parker & Phillips, 2017; Van Dyke & McElree, 2011). However, the application of differential cue weighting in retrieval for sentence processing remains poorly understood. An important task for future work is to investigate how both cue weighting and prediction impact cue combinatorics in retrieval for language processing.

6. Conclusion

This paper addressed the question of how different types of cues are combined to access linguistic information in memory. This study directly compared linear and nonlinear cue combination rules using predictions derived from computational modeling. Those predictions were then tested with antecedent-reflexive dependencies using a self-paced reading

design that manipulated the degree of match between the target antecedent and reflexive. Results showed that target items that matched all the cues (full match) were favored more than target items with a 1- and 2-feature mismatch and that the 1- and 2-feature mismatches slowed reading times in comparable amounts. These results are consistent with the predictions of a nonlinear rule and provide evidence against models that assume that retrieval for dependency formation always utilizes a linear combination. A follow-up set of simulations in Experiment 3 showed that a nonlinear rule also captures previous demonstrations of interference effects in antecedent retrieval for reflexive licensing, lending additional support to the current proposal that retrieval for reflexive licensing utilizes a nonlinear cue combination method. These results shed new light on our understanding of the factors that determine the success and failure of retrieval in language processing. Specifically, the results are informative about how different types of cues combine at the retrieval site and how the method of cue combination impacts the accessibility of linguistic information in memory—a topic that has received little attention in previous research.

Notes

1. These predictions assume that no feature-matching distractor is present. Experiment 3 outlines the predictions for contexts with a distractor.
2. A previous version of Experiment 2 presented at the 2017 CUNY Human Sentence Processing Conference tested three violations. The third violation involved a 3-feature mismatch using the first-person plural pronoun *we* (e.g., *We . . . himself*). The results of this experiment, which are provided in Data S1, are fully consistent with the current results and theoretical conclusions. This additional condition introduces several challenges due to the limitations of English morphology, motivating a new version of the experiment reported here. In particular, *we* is an unheralded pronoun that lacks a referent, it differs in length relative to the other target NPs, and it is not clear that the underspecification or lack of a gender feature triggers a cue mismatch penalty for the reflexive.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychology Review*, *111*, 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036.

Anderson, J. R., Budiu, R., & Reder, L. M. (2001). A theory of sentence memory as part of a general theory of memory. *Journal of Memory and Language*, *45*, 337–367. https://doi.org/10.1006/jmla.2000.2770.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychology Review*, *96*, 703–719. https://doi.org/10.1037/0033-295X.96.4.703.

Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

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*, 748–769. https://doi.org/10.1037/0278-7393.28.4.748.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. Available at http://CRAN.R-project.org/package=lme4.

Boston, M. F., Hale, J., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*, 301–349. https://doi.org/10.1080/01690965.2010.492228.

Bothell, D. (2007). ACT-R 6.0 reference manual. Available at http://act-r.psy.cmu.edu/wordpress/wp-content/themes/ACT-R/actr6/reference-manual.pdf.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological)*, *26*, 211–252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review*, *20*, 243–268. https://doi.org/10.3758/s13423-012-0369-9.

Chen, Z., Jäger, L., & 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 (studies in generative grammar)* (pp. 43–62). Berlin: Mouton de Gruyter.

Chomsky, N. (1981). *Lectures on government and binding*. Berlin: Mounten de Gruyter.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. Duncan Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York, NY: Wiley.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37–60. https://doi.org/10.3758/BF03210740.

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

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*, 87–185. https://doi.org/10.1017/S0140525X01003922.

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*, 188–219. https://doi.org/10.1080/01690965.2010.548391.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory & Language*, *75*, 117–139. https://doi.org/10.1016/j.jml.2014.05.006.

Dillon, B. (2014). Syntactic memory in comprehension of reflexive dependencies: An overview. *Language and Linguistics Compass*, *8*, 171–187. https://doi.org/10.1111/lnc3.12075.

Dillon, B., Chow, W. Y., Wagers, M., Guo, T., Liu, F., & Phillips, C. (2016). The structure-sensitivity of search: Evidence from Mandarin Chinese. *Frontiers in Psychology*, *5*, 1–16.

Dillon, B., Chow, W. Y., & Xiang, M. (2016). The relationship between anaphor features and antecedent retrieval: Comparing Mandarin Ziji and Ta-Ziji. *Frontiers in Psychology*, *6*, 1–16. https://doi.org/10.3389/fpsyg.2015.01966.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Lanuage*, *69*, 85–103. https://doi.org/10.1016/j.jml.2013.04.003.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language*. New York: McGraw-Hill.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*, 357–383. https://doi.org/10.1016/j.jml.2006.07.004.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*, 291–325. https://doi.org/10.1016/0010-0277(78)90002-1.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Gibson, E., Paintadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science*, *24*, 1079–1088. https://doi.org/10.1177/0956797612463705.

Gillund, G., & Shiffrin, R. M. (1984). A retreival model for both recognition and recall. *Psychological Review*, *91*, 1–67. https://doi.org/10.1037/0033-295X.91.1.1.

Grus, J. (2015). *Data science from scratch*. Sebastopol, CA: O'Reilly.

Halle, M., & Stevens, K. (1959). Analysis by synthesis. In W. Wathen-Dunn & L. E. Woods (Eds.), *Proceedings of the seminar on speech compression and processing* (pp. 1–4). Bedford, MA: L.G.Hanscom Field.

Halle, M., & Stevens, K. (1963). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory*, *8*, 155–159. https://doi.org/10.1109/TIT.1962.1057686.

Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments, and Computers*, *16*, 96–101. https://doi.org/10.3758/BF03202365.

Jäger, L., Benz, L., Roeser, J., Dillon, B., & Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. *Frontiers in Psychology*, *6*, 1–18. https://doi.org/10.3389/fpsyg.2015.00506.

Jäger, L., Engelmann, F., & Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, *6*, 1–24. https://doi.org/10.3389/fpsyg.2015.00617.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.004.

Jarvella, R. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, *10*, 409–416. https://doi.org/10.1016/S0022-5371(71)80040-3.

Jarvella, R., & Herman, S. (1972). Clause structure of sentences and speech processing. *Perception and Psychophysics*, *11*, 381–384. https://doi.org/10.3758/BF03206272.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193–224. https://doi.org/10.1146/annurev.psych.59.103006.093615.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*, 15–47. https://doi.org/10.1016/0010-0277(72)90028-5.

Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (Ed.), *Syntax and semantics* (pp. 155–179). New York: Academic Press.

Kohonen, T. (1980). *Content-addressable memories*. Berlin: Springer-Verlag. https://doi.org/10.1007/978-3-642-96552-4.

Kush, D. (2013). Respecting relations: Memory access and antecedent retrieval in incremental sentence processing. Ph.D. Dissertation, University of Maryland.

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an SOV language: Implications for retrieval strategies. *Frontiers in Psychology*, *5*, 1–12. https://doi.org/10.3389/fpsyg.2014.01252.

Lago, S., Shalom, D., Sigman, M., Lau, E., & Phillips, C. (2015). Agreement processes in Spanish comprehension. *Journal of Memory and Language*, *82*, 133–149. https://doi.org/10.1016/j.jml.2015.02.002.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In D. Yarowsky, et al. (Eds.), *Proceedings of the 13th conference on empirical methods in natural language processing* (pp. 234–243). Stroudsburg, PA: Association for Computational Linguistics.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, *25*, 93–115. https://doi.org/10.1007/BF01708421.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational Rationality: Linking Mechanisms and Behavior Through Bounded Utility Maximization. *Topics in Cognitive Science*, *6*, 279–311. https://doi.org/10.1111/tops.12086.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375–419. https://doi.org/10.1207/s15516709cog0000_25.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, *10*, 447–454. https://doi.org/10.1016/j.tics.2006.08.007.

Martin, A. E. (2016). Langauge processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology*, *7*, 1–17. https://doi.org/10.3389/fpsyg.2016.00120.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, *58*, 879–906. https://doi.org/10.1016/j.jml.2007.06.010.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1231–1239. https://doi.org/10.1037/a0016271.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from Sluicing. *Journal of Memory and Language*, *64*, 327–343. https://doi.org/10.1016/j.jml.2010.12.006.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage*, *59*, 1859–1869. https://doi.org/10.1016/j.neuroimage.2011.08.057.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*, 155–200. https://doi.org/10.1023/A:1005184709695; https://doi.org/10.1023/a:1005140927442.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation—Advances in research and theory* (pp. 155–200). San Diego, CA: Academic Press.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Pscyhology*, *18*, 346–373. https://doi.org/10.1037/0096-3445.118.4.346.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*, 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *101*, 343–352. https://doi.org/10.1037/0033-295X.101.2.343.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. Duncan Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). New York: Wiley.

Miller, G. A., & Isard, S. D. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, *2*, 217–228. https://doi.org/10.1016/S0022-5371(63)80087-0.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269. https://doi.org/10.3758/BF03213879.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81. https://doi.org/10.1146/annurev.psych.53.100901.135131.

Nicenboim, B., Logačev, P., Gattei, C., & Vasishth, S. (2016). When high-capacity readers slow down and low-capacity readers speed up: Working memory and locality effects. *Frontiers in Psychology*, *7*, 1–24. https://doi.org/10.3389/fpsyg.2016.00280.

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34. https://doi.org/10.1016/j.jml.2017.08.004.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 5–19. https://doi.org/10.1007/BF01069043.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872. https://doi.org/10.1016/j.jesp.2009.03.009.

Paape, D. (2016). Filling the silence: Reactivation, not reconstruction. *Frontiers in Psychology*, *7*, 1–18. https://doi.org/10.3389/fpsyg.2016.00027.

Parker, D., Lago, S., & Phillips, C. (2015). Interference in the processing of adjunct control. *Frontiers in Psychology*, *6*, 1–13. https://doi.org/10.3389/fpsyg.2015.01346.

Parker, D., & Lantz, D. (2017). Encoding and accessing linguistic representations in a dynamically structured holographic memory system. *Topics in Cognitive Science*, *9*, 51–68. https://doi.org/10.1111/tops.12246.

Parker, D., & Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition*, *157*, 321–339. https://doi.org/10.1016/j.cognition.2016.08.016.

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language*, *94*, 272–290. https://doi.org/10.1016/j.jml.2017.01.002.

Patil, U., Vasishth, S., & Lewis, R. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology*, *7*, 1–18. https://doi.org/10.3389/fpsyg.2016.00329.

Patson, N. D., & Husband, M. (2015). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, *69*, 950–971. https://doi.org/10.1080/17470218.2014.992445.

Pearlmutter, N., Garnsey, S., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456. https://doi.org/10.1006/jmla.1999.2653.

Poeppel, D., & Monahan, P. (2010). Feedforward and feedback in speech perception: Revisiting analysis-by-synthesis. *Language and Cognitive Processes*, *26*, 935–951. https://doi.org/10.1080/01690965.2010.493301.

R Development Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Raajimakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134. https://doi.org/10.1037/0033-295X.88.2.93.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510. https://doi.org/10.1037/0033-2909.114.3.510.

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, *181*, 574–576. https://doi.org/10.1126/science.181.4099.574.

Reinhart, T. (1976). The syntactic domain of anaphora. Ph.D. dissertation, Massachusetts Institute of Technology.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, *24*, 657–720.

Runner, J., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrases. *Cognitive Science*, *30*, 193–241. https://doi.org/10.1207/s15516709cog0000_58.

Slioussar, N., & Malko, A. (2016). Gender agreement attraction in Russian: Production and comprehension evidence. *Frontiers in Psychology*, *7*, 1–20. https://doi.org/10.3389/fpsyg.2016.01651.

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, *60*, 308–327. https://doi.org/10.1016/j.jml.2008.11.002.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition*, *114*, 447–454. https://doi.org/10.1016/j.cognition.2009.11.003.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*, 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215. https://doi.org/10.1016/j.jml.2014.07.003.

Trommershäuser, J., Körding, K., & Landy, M. S. (2011). *Sensory cue integration*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195387247.001.0001.

Tucker, M. A., & Almeida, D. (2017). The complex structure of agreement errors: Evidence from distributional analyses of agreement attraction in Arabic. In A. Lamont & K. Tetzloff (Eds.), *Proceedings of the 47th Meeting of the North East Linguistics Society* (pp. 45–54). Amherst, MA: GLSA.

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*, 1–21. https://doi.org/10.3389/fpsyg.2015.00347.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 407–430. https://doi.org/10.1037/0278-7393.33.2.407.

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 misanalyzed ambiguities. *Journal of Memory and Language*, *49*, 285–316. https://doi.org/10.1016/S0749-596X(03)00081-0.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*, 157–166. https://doi.org/10.1016/j.jml.2006.03.007.

Van Dyke, J. A., & McElree, B. (2007). Similarity-based proactive and retroactive interference reduces quality of linguistic representations. Poster presented at the CUNY Conference on Human Sentence Processing. San Diego, CA.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*, 247–263. https://doi.org/10.1016/j.jml.2011.05.002.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*, 685–712. https://doi.org/10.1080/03640210802066865.

Vasishth, S., & Nicenboim, B. (2016). Statistical Methods for Linguistic Research: Foundational Ideas - Part 1. *Language and Linguistics Compass*, *10*, 349–369.

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

Wagers, M. W. (2008). The structure of memory meets memory for structure in linguistic cognition. Ph.D. Dissertation, University of Maryland.

Wundt, W. (1904). The psychology of the sentence. Reprinted in A. L. Blumenthal (Ed.). (1970). *Language and psychology: Historical aspects of psycholinguistics*. (pp. 9–32). New York: Wiley.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, *108*, 40–55.

Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology*, *4*, 1–19.

Supporting Information

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

Data S1. Supplementary materials.