Interference in Language Processing Reflects Direct-Access Memory Retrieval: Evidence from Drift-Diffusion Modeling

Dan Parker (dparker@wm.edu)
Department of English, Linguistics Program, College of William & Mary

Adam An (zan@email.wm.edu)
Linguistics Program, College of William & Mary

Abstract

Many studies on memory retrieval in language processing have identified similarity-based interference as a key determinant of comprehension. The broad consensus is that similarity-based interference reflects erroneous retrieval of a non-target item that matches some of the retrieval cues. However, the mechanisms responsible for such effects remain debated. Activation-based models of retrieval (e.g., Lewis & Vasishth, 2005) claim that any differences in processing difficulty due to interference in standard RT measures and judgments reflect differences in the speed of retrieval (i.e., the amount of time it takes to retrieve a memory item). But this claim is inconsistent with empirical data showing that retrieval time is constant due to the use of a direct-access procedure (e.g., McElree, 2000, 2006). According to direct-access accounts, differences in judgments or RTs due to interference arise from differences in the quality or availability of the candidate memory representations, rather than differences in retrieval speed. To adjudicate between these accounts, we employed a novel methodology that combined a high-powered (N = 200) two-alternative forced-choice study on interference effects with drift diffusion modeling to disassociate the effects of retrieval speed and representation quality. Results showed that the presence of a distractor that matched some of the retrieval cues lowered asymptotic accuracy, reflecting an effect of representation quality, but did not affect retrieval speed, consistent with a direct-access procedure. These results suggest that the differences observed in RTs and judgment studies reflect differences in the ease of integrating the retrieved item back into the current processing stream, rather than differences in retrieval speed.

Keywords: language processing; working memory; interference; two-alternative forced-choice task; drift diffusion modeling

Introduction

Successful language comprehension requires the ability to encode complex linguistic representations in memory and accurately access specific pieces of information in those representations to guide further elaboration of the discourse. For example, to relate the verb play in (1) with its subject for number agreement and thematic binding, memory retrieval mechanisms must access the encoding of the plural target subject kids and ignore featurally-similar items in non-target positions, such as the embedded plural noun teachers.

(1) The kids[pl] [that the teachers[pl] watched closely] played on the slide.

However, many studies have shown that featurally-similar items in non-target positions can interfere with retrieval of the target, impacting judgments of acceptability and reading times (for a review, see Parker, Shvartsman, & Van Dyke, 2017). Such effects are commonly called “similarity-based interference” (Gordon, Hendrick, & Johnson, 2001; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke, 2007; Van Dyke & Johns, 2012; Van Dyke & McElree, 2006, 2011). The goal of the current study is to help identify the source of such effects in language comprehension.

Often, interference from non-target items during retrieval for linguistic dependency formation slows reading times and lowers acceptability. This type of interference is called “inhibitory” interference (see Jäger, Engelmann, & Vasishth, 2017, for a review) and occurs in multiple match configurations where the target and a distractor overlap in some features that are relevant for retrieval, as in (1).

It has also been shown that interference can sometimes speed up processing and boost acceptability, resulting in an effect known as “facilitatory interference” or more commonly, “attraction” (Jäger et al., 2017). Attraction arises when the target and distractor are distinct in feature content, but neither is a perfect match to the retrieval cues. Such effects are commonly observed in the processing of subject-verb number agreement. For instance, Wagers and colleagues (2009) examined the comprehension of subject-verb agreement in sentences like (2) using self-paced reading and speeded acceptability judgments. The sentences in (2c-d) are ungrammatical because the plural verb were does not agree in number with the head of its subject noun phrase (NP) key.

(2) a. The key to the cabinets certainly was rusty …
   b. The key to the cabinet certainly was rusty …
   c. *The key to the cabinets certainly were rusty …
   d. *The key to the cabinet certainly were rusty …

Wagers and colleagues found that in grammatical sentences like (2b), the number marking on the plural attractor cabinet(s) did not impact acceptability or RTs after the verb. However, in ungrammatical sentences like (2c), the plural distractor cabinets (the “attractor”), which matched the number of the verb were, boosted acceptability and facilitated RTs after the verb, relative to the ungrammatical condition with the singular noun cabinet (2d). Wagers and colleagues argued that the effects of facilitation and boosted
acceptability of sentences like (2c) were due to erroneous retrieval of the plural attractor. According to their account, retrieval functions as an error-driven repair mechanism that is triggered by the detection of an agreement violation. In (2), the subject NP predicts the number of the verb. When the verb form violates this prediction, as in (2c-d), the parser engages a cue-based retrieval at the verb to recover a number matching noun to license agreement. The attractor cabinets in (2c) will sometimes be incorrectly retrieved because it matches the verb in number, easing processing in a way that facilitates reading and boosts overall acceptability. In the grammatical conditions (2a-b), the verb fulfills the number prediction made by the subject NP, and therefore retrieval is not engaged, reducing the likelihood of attraction.

Alternative accounts exist, but many researchers concur that agreement attraction arises due to incorrect memory retrieval (e.g., Dillon, Mishler, Sloggett, & Phillips, 2013; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Phillips, Wagers, & Lau, 2011; Schlueter, Williams, & Lau, 2018; Tanner, Nicol, & Brehm, 2014; Tucker & Almeida, 2017; Tucker, Idrissi, & Almeida, 2015). However, the reason for why incorrect retrieval facilitates RTs is debated and the relationship between RTs and retrieval accuracy remains underspecified.

For example, the prominent activation-based model of memory retrieval (ACT-R) developed by Lewis and Vasishth (Lewis & Vasishth, 2005) claims that the differences in RTs due to facilitatory interference (e.g., 2c vs. 2d) reflect differences in the speed of retrieval (i.e., the amount of time it takes to retrieve a memory item). In their model, the strength of an item’s activation at the moment of retrieval determines the item’s retrieval accuracy and its retrieval speed, such that items with higher activation are more likely to be retrieved and will be retrieved more quickly than items with a lower activation. In sentences that show attraction, like (2c), the plural attractor will have a higher activation than the singular attractor in (2d) because it provides a better match to the cues of the verb, and therefore will have a faster retrieval latency, resulting in faster RTs and boosted acceptability.

The activation-based model has been shown to provide a good fit to a wide range of behavioral data (Parker et al., 2017), but it is inconsistent with empirical evidence showing that retrieval speed is constant (i.e., time invariant) due to the use of a direct-access procedure (Martin & McElree, 2008, 2009, 2011; McElree, 2000; McElree & Dosher, 1989; McElree, Foraker, & Dyer, 2003; Van Dyke & McElree, 2011). According to direct-access accounts, the cues at retrieval make direct contact with the items in memory based on their content, rather than their location, which allows items to be retrieved at a constant speed, regardless of their position or dependency length. Items are differentially activated based on their (partial) match to the cues and the item that is most strongly activated is retrieved for dependency formation. On this view, the differences in RTs in (2c) vs. (2d) reflect differences in the quality (activation strength or availability) of the candidate memory representations, rather than differences in retrieval speed. For instance, the attractors in (2c) and (2d) will be retrieved in equal time, but the plural attractor in (2c) will be integrated into the processing stream more quickly because it provides a better match to the cues, resulting in faster RTs and boosted acceptability.

At present, it is difficult to distinguish between these accounts because the typical measures used to investigate attraction (e.g., reading times and judgments) do not discriminate between effects that arise from differences in retrieval speed and differences in representation quality. Furthermore, the argument for direct-access is based entirely on studies of inhibitory interference where distractors slow RTs (see Parker et al., 2017, for a review) and it remains unclear whether facilitatory interference effects like attraction show the same retrieval dynamics as inhibitory interference. These issues are addressed in the present study.

The Present Study

The goal of the present study is to tease apart existing predictions about retrieval speed and representation quality to better understand the source of facilitatory interference effects in language processing. Previously, research on retrieval in sentence processing has relied on the speed-accuracy trade-off (SAT) procedure (Dosher, 1979; Reed, 1973; Wickelgren, 1977) to examine the effects of retrieval speed orthogonally from effects of representation quality. In an SAT task, participants read sentences presented via rapid serial visual presentation (RSVP) and make binary judgments about sentence acceptability at cue intervals, ranging from before the tail of the critical dependency to 3-6 seconds after the dependent constituent is presented. Participants’ average performance at these cue times is interpolated into an exponential curve that summarizes the speed-accuracy tradeoff function revealing the time course of retrieval. Importantly, by sampling a range of intervals, independent estimates of retrieval speed and accuracy become available. This method provides a profile of memory retrieval processes that is characterized by three parameters: (i) the asymptote, which reflects retrieval accuracy, (ii) the intercept, which reflects the time to retrieve an item from memory, and (iii) rate, which reflects the speed at which accuracy grows from the intercept to the asymptote. Differences in either the intercept or rate are presented as evidence for differences in retrieval speed, and differences in asymptote are taken to reflect differences in representation quality.

The SAT methodology has been pivotal in arguing that retrieval for sentence processing employs a time-invariant direct-access procedure (e.g., Martin & McElree, 2008, 2009, 2011; McElree, 2000; McElree & Dosher, 1989; McElree, Foraker, & Dyer, 2003; Van Dyke & McElree, 2011). According to direct-access accounts, the cues at retrieval make direct contact with the items in memory based on their content, rather than their location, which allows items to be retrieved at a constant speed, regardless of their position or dependency length. Items are differentially activated based on their (partial) match to the cues and the item that is most strongly activated is retrieved for dependency formation. On this view, the differences in RTs in (2c) vs. (2d) reflect differences in the quality (activation strength or availability) of the candidate memory representations, rather than differences in retrieval speed. For instance, the attractors in
also been used to jointly analyze the effects of accuracy and processing speed and model the timing of retrieval (Chen & Husband, 2018; McElree & Dosher, 1989; Ratcliff, 1978; Ratcliff, Smith, Brown, & McKoon, 2016). Importantly, recent research on memory retrieval in sentence processing has shown that DDM yields results that are comparable to the more costly SAT methodology (Chen & Husband, 2018). Based on these results, we extended the DDM methodology to test existing predictions about retrieval speed and representation quality regarding facilitatory interference effects (i.e., activation-based vs. direct-access models of retrieval).

DDM uses data from two-alternative forced choice (2AFC) tasks to generate a conditional cumulative distribution function (CDF) that relates a time \( T \) to the probability that a correct response is faster than or equal to \( T \). Crucially, it relies on four parameters that have been argued to reflect distinct underlying memory retrieval processes in sentence processing (Chen & Husband, 2018):

(i) \( \tau \) non-decision time: encoding and motor response time, including the time to extract the relevant information from memory to make a decision

(ii) \( \alpha \) boundary separation: the amount of evidence needed to make a decision

(iii) \( \delta \) drift rate: rate of evidence accumulation

(iv) \( \beta \) response bias: the bias to respond to a particular alternative

In the current study, we tested a standard agreement attraction paradigm like that in (2) as a hallmark of facilitatory interference in a high-powered (N=200) 2AFC experiment and modeled the data using drift diffusion modeling to distinguish between effects arising from differences in retrieval speed vs. differences in representation quality. Recent research has used DDM to investigate how response biases impact the amount of attraction in sentences like (2c), as measured with the \( \beta \) response bias parameter (Hammerly, Staub, & Dillon, unpublished ms.). However, this work did not test the current predictions about retrieval speed, nor did it explicitly address the question of retrieval time. The present study applies the same methodology, but focuses on the issue of processing dynamics to better understand why interference eases processes in sentences like (2c).

Under both the activation-based and direct-access accounts, facilitatory interference should negatively impact asymptotic accuracy (DDM \( \delta \) drift rate), such that the sentences that give rise to attraction (2c) should have an overall lower accuracy relative to the other conditions (2a, b, d). Where the accounts differ, however, is in their predictions for processing dynamics. If facilitatory interference arises due to faster memory access, as claimed by activation-based accounts, then we should see a faster intercept (\( \tau \) non-decision time) for sentences that show attraction (2c). By contrast, if retrieval occurs via direct access, then the intercept parameters should be comparable across conditions.

### Method

#### Participants

Participants were 200 college-age native speakers of English. The large sample size was chosen to ensure high statistical power (i.e., reduce Type II error) and accurate estimation of the DDM parameters. All participants provided informed consent and received credit in an introductory psychology or linguistics course. All participants were naïve to the purpose of the experiment. The experiment lasted approximately 20 mins.

#### Materials

Experimental materials consisted of 64 sets of 4 items like those shown in Table 1. The high number of item sets was chosen to ensure a stable estimation of the DDM parameters. Experimental conditions consisted of a \( 2 \times 2 \) factorial design that crossed grammaticality (grammatical/ungrammatical) and attractor number (singular/plural). In all conditions, the subject head noun was modified by a prepositional phrase that contained the attractor. The critical verb was always a full lexical verb in sentence-final position. An adverb created a buffer between the subject and the critical verb to control for processing effects associated with plural nouns (see Wagers et al, 2009). Grammaticality was manipulated by varying the verb number such that it either matched or mismatched the number of the subject head noun. Attractor number was manipulated by varying the number of the attractor such that it either matched or mismatched the verb number.

The 64 target items were distributed across 4 lists in a Latin square design and combined with 66 fillers. Half of the fillers were ungrammatical, yielding an overall grammatical-to-ungrammatical ratio of 1:1. Approximately half of the grammatical fillers involved sentence-final plural verbs in structures similar to the target items and approximately half of the ungrammatical fillers involved sentence-final singular verbs to unconfound grammaticality with verb number in the target items. The remaining fillers involved relative clause structures from an unrelated experiment.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical PL attractor</td>
<td>The tutor for the students often rambles.</td>
</tr>
<tr>
<td>Grammatical SG attractor</td>
<td>The tutor for the students often rambles.</td>
</tr>
<tr>
<td>Ungrammatical PL attractor</td>
<td>The tutor for the students often rambles.</td>
</tr>
<tr>
<td>Ungrammatical SG attractor</td>
<td>The tutor for the student often rambles.</td>
</tr>
</tbody>
</table>
Procedure

Sentences were presented using Ibex (Drummond) one word at a time in the center of the screen in RSVP mode with a stimulus onset asynchrony (SOA) of 300 ms per word and an interstimulus interval (ISI) of 100 ms. Participants were instructed to read each sentence carefully and judge whether each sentence was an acceptable sentence of English. A response screen appeared for 3 s at the end of each sentence during which participants made a ‘yes/no’ response by button press. If participants waited longer than 3 s to respond, they were given feedback that their response was too slow. The order of presentation was randomized for each participant.

Data Analysis

All data were included in the analyses. A logistic mixed-effects model was fit to the judgment accuracy data and a linear model was fit to the raw response latencies using the lme4 package (Kuznetsova, Brockhoff, & Christensen, 2014) in the R software environment (R Development Core Team, 2018), with fixed factors for the experimental manipulations (i.e., grammaticality and attractor number) and their interaction. All models were fit with the maximal random effects structure supported by the data (Barr, Levy, Scheepers, & Tily, 2013). An effect was considered significant if $|t/z| > 2$.

For the DDM analysis, the RWeiner package (Wabersich & Vandekerckhove, 2014) was used to fit a Weiner drift diffusion model to each condition for each participant. Parameter values that did not converge were excluded, following Chen and Husband (2018). A linear model was fit to the by-participant parameter fits following the same procedure used in the analysis of the response latencies. All data and code are available via Open Science Framework: https://osf.io/bu2kh/.

Results

Judgments and Response Latencies

Figure 1 shows the percentage of ‘yes’ responses and latencies (in ms) for the four experimental conditions. Main effects of grammaticality and attractor were observed in the judgments and latencies ($z > 3$ in all cases). Grammatical sentences were more likely to be accepted and had faster latencies than ungrammatical sentences, and sentences with a plural attractor were more likely to be accepted and had longer latencies than sentences with a singular attractor. Crucially, judgments also showed a significant interaction of grammaticality with attractor number ($z = -12.48$). Planned pairwise comparisons revealed that this interaction was carried by the ungrammatical conditions: participants were more likely to accept an ungrammatical sentence when a plural attractor was present ($z = 12.11$). No such effect was observed in the grammatical conditions ($z = -1.13$). This profile reflects the behavioral signature of agreement attraction (Phillips et al., 2011) and provides an appropriate basis to examine the relationship between retrieval accuracy and retrieval speed using the DDM methodology.

Drift Diffusion Model (DDM)

Average DDM parameters by condition are shown in Table 2, and the $t$-values for model estimates of effects on DDM parameters are shown in Table 3. Figure 2 shows the cumulative density of accurate responses as a function of response time by condition. DDM revealed an effect of attraction on $\delta$ drift rate (asymptotic accuracy), qualified by an interaction between grammaticality and attractor number, such that participants were less accurate in ungrammatical sentences with a plural attractor than in those with a singular attractor. This effect is predicted by both accounts.

With respect to processing dynamics, which is where the accounts diverge, DDM revealed no significant effect of attraction on the processing dynamics reflected in $\tau$ non-decision time (intercept). These results suggest that agreement attraction impacts retrieval accuracy but not retrieval speed, consistent with a direct-access model of memory retrieval.

Results also showed a main effect of grammaticality on $\tau$ non-decision time (intercept), as grammatical sentences showed faster response latencies than ungrammatical sentences. This effect is unrelated to interference and likely reflects facilitation due to predictive processing in the grammatical conditions (Wagers et al., 2009).
Table 2: DDM parameters by condition.

<table>
<thead>
<tr>
<th></th>
<th>( \tau )</th>
<th>( \alpha )</th>
<th>( \delta )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical PL attr</td>
<td>0.26</td>
<td>2.03</td>
<td>0.90</td>
<td>0.60</td>
</tr>
<tr>
<td>Grammatical SG attr</td>
<td>0.23</td>
<td>2.18</td>
<td>1.12</td>
<td>0.62</td>
</tr>
<tr>
<td>Ungrammatical PL attr</td>
<td>0.30</td>
<td>1.99</td>
<td>-0.24</td>
<td>0.47</td>
</tr>
<tr>
<td>Ungrammatical SG attr</td>
<td>0.29</td>
<td>2.39</td>
<td>-1.05</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Figure 2: DDM estimations of the cumulative density of “yes” (1) and “no” (0) responses as a function of response time by condition. PL = plural, SG = singular.

Discussion

The goal of the present study was to distinguish between existing predictions about retrieval speed and retrieval accuracy to better understand the source of interference effects in language processing. On the one hand, activation-based models of retrieval (e.g., Lewis & Vasishth, 2005) claim that differences in processing difficulty due to interference in standard RT measures and judgments reflect differences in the speed of retrieval. On the other hand, direct-access accounts claim that differences in judgments or RTs due to interference arise from differences in the quality of the candidate memory representations, rather than differences in retrieval speed, based on behavioral data showing that retrieval time is constant. To adjudicate between these accounts, we tested for facilitatory interference paradigm in a high-powered 2AFC experiment and modeled the results using DDM to disassociate the effects retrieval speed and representational quality.

Results of the 2AFC task replicated the classic attraction profile, such that ungrammatical sentences with a plural attractor that matched the number of the verb showed boosted acceptability relative to ungrammatical sentences with a singular attractor. Results of the DDM analysis revealed that in the ungrammatical conditions, the presence of a number-matching plural attractor lowered overall asymptotic accuracy, but did not affect retrieval speed.

The lack of an effect on non-decision time is consistent with the predictions of a direct-access procedure. These results suggest that the differences in judgments and RTs observed in agreement attraction studies reflect differences in the ease of integrating the retrieved item back into the current processing stream, rather than differences in retrieval speed.

More specifically, we argue that the quality of the memory representation (described in terms of activation strength) impacts the post-access stage of “binding”, rather than the speed of access. In the memory literature, binding refers to the mechanisms by which information in memory is integrated together (Cohen & Eichenbaum, 1993; Hagoort, 2003; van der Velde & de Kamps, 2006), and it has been suggested that the effort required for integration is governed, in part, by the item’s representation quality (Budiu & Anderson, 2004). On this view, retrieval of an item that satisfies at least some of the search criteria, such as the number matching attractor in sentences like (2c), will make post-retrieval integration faster compared to integration of an item that does not satisfy the search requirements, such as in (2d), giving rise to facilitatory interference.

More broadly, the current results are consistent with the recent claim that differences in the quality or availability of the information in memory leads to differences in accuracy and that those differences underlie the differences in reaction time studies (Martin & McElree, 2018). The current study extends this conclusion to facilitatory interference, motivating a unified analysis of inhibitory and facilitatory interference as the signature of direct-access retrieval.

Table 3: \( t \)-values for linear mixed effects model estimates on DDM parameters with 95% Cis in brackets.

<table>
<thead>
<tr>
<th></th>
<th>( \tau )</th>
<th>( \alpha )</th>
<th>( \delta )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammaticality</td>
<td>-4.85 [-0.08, 0.31]</td>
<td>-1.77 [-0.45, 0.02]</td>
<td>21.92 [1.98, 2.37]</td>
<td>11.85 [0.17, 0.24]</td>
</tr>
<tr>
<td>Attractor number</td>
<td>1.17 [-0.00, 0.03]</td>
<td>-3.64 [-0.62, -0.18]</td>
<td>11.71 [0.67, 0.94]</td>
<td>2.55 [0.03, 0.09]</td>
</tr>
<tr>
<td>Interaction</td>
<td>1.04 [-0.01, 0.04]</td>
<td>1.88 [-0.01, 0.53]</td>
<td>-11.47 [-1.20, -0.85]</td>
<td>-4.28 [-0.12, -0.04]</td>
</tr>
</tbody>
</table>
References


