frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 04 February 2026
DOI 10.3389/flang.2025.1708378

Check for updates

OPEN ACCESS

EDITED BY
Giosuè Baggio,
Norwegian University of Science and Technology, Norway

REVIEWED BY
John E. Drury,
Brooklyn College (CUNY), United States
Theodora Alexopoulou,
University of Cambridge, United Kingdom

*CORRESPONDENCE
Maayan Keshev
✉ maayan.keshev@mail.huji.ac.il

RECEIVED 18 September 2025
REVISED 18 December 2025
ACCEPTED 31 December 2025
PUBLISHED 04 February 2026

CITATION
Keshev M, Cartner M, Meltzer-Asscher A and Dillon B (2026) Representation distortion contributes to agreement attraction in comprehension.
Front. Lang. Sci. 4:1708378.
doi: 10.3389/flang.2025.1708378

Representation distortion contributes to agreement attraction in comprehension

Maayan Keshev[1]*, Mandy Cartner[2], Aya Meltzer-Asscher[2,3] and Brian Dillon[4]

[1]Department of Linguistics, Hebrew University of Jerusalem, Jerusalem, Israel, [2]Department of Linguistics, Tel Aviv University, Tel Aviv-Yafo, Israel, [3]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv-Yafo, Israel, [4]Department of Linguistics, University of Massachusetts Amherst, Amherst, MA, United States

Sentence comprehension relies on encoding linguistic items in memory and accessing them subsequently to form linguistic dependencies. This makes processing susceptible to memory interference. Interference, such as the distortion of memory representations or access to irrelevant memory items, can lead to misinterpretation or grammatical errors. Over the years, research on agreement attraction has debated whether this hallmark of memory interference reflects limits of the retrieval mechanism, or inaccuracy of the encoded representations that retrieval targets. We present some evidence in favor of representational accounts of memory interference. Our findings include partial evidence for three kinds of representational effects: (a) the ungrammaticality illusion, a pattern by which attraction arises without misleading retrieval cues; (b) number errors rather than noun errors in final interpretation; and (c) mitigation of attraction when additional markers of the subject's number are available, which we label feature updating. Together, the findings seem to suggest that feature distortion in the content of memory representations contributes to attraction effects. We propose that models of memory mechanisms that mediate dependency formation should incorporate malleable representations rather than stable ones.

KEYWORDS

agreement attraction, encoding, feature distortion, grammaticality illusion, sentence processing

1 Introduction

To understand a sentence, a reader or a listener has to integrate words that might be far apart. Such integration processes require encoding words into memory and subsequently accessing their memory encodings when they are necessary for interpretation. Research into these memory processes has closely examined grammatical illusions—cases where comprehenders temporarily form ungrammatical associations between two elements of a larger linguistic expression (Lewis and Phillips, 2015; Phillips et al., 2011). It has been suggested that such illusions reflect access to a structurally irrelevant item in memory, i.e., retrieval interference (Cunnings and Sturt, 2018; Jäger et al., 2017; Vasishth et al., 2008; Wagers et al., 2009; among others). This research has mostly assumed that memory encodings provide stable, veridical representations of the input, at least as a simplifying assumption (following Lewis and Vasishth, 2005). However, mounting evidence suggests that illusions can also arise from the distortion of the feature content of a given item in memory, yielding malleable, potentially non-veridical representations of linguistic input in working memory (Brehm et al., 2021; Eberhard et al., 2005; Hammerly et al., 2019; Keshev et al., 2025; Keshev and Meltzer-Asscher, 2024; Laurinavichyute and von der Malsburg, 2024; Paape et al., 2021; Staub, 2009; Yadav et al., 2022).

Here, we further explore (across four experiments) the ways in which memory encoding mechanisms are vulnerable to interference that distorts representations. We also ask whether additional cues to the feature content of items in memory can serve to update those representations as the sentence unfolds (in line with Keshev and Meltzer-Asscher, 2024; Molinaro et al., 2008). We investigate this in the context of agreement attraction. We argue that inaccurate attribution of agreement features to items in memory interferes with agreement processing. This suggests that understanding how feature-item mapping is encoded and maintained in working memory is crucial for modeling sentence processing (as proposed by Keshev et al., 2025).

1.1 Agreement attraction: representational distortion and retrieval interference

Errors in the formation of subject-verb agreement have been informative about the memory mechanisms that subserve sentence processing. Agreement errors can arise when a non-subject noun (a *distractor*) differs in agreement features from the *target* phrase that should control agreement, typically the subject phrase. This type of error was first identified in production, where it was found that in a configuration like (1a), speakers sometimes produce a verb that agrees with the structurally irrelevant distractor (i.e., "were"). Such errors arise in a non-negligible proportion of cases (∼20%), as compared against a baseline where the distractor matches the subject (1b) (Bock and Cutting, 1992; Bock and Eberhard, 1993; Bock and Miller, 1991; Franck et al., 2002; Hartsuiker et al., 2003; Haskell et al., 2010; Slioussar, 2018; Vigliocco et al., 1995). This phenomenon is referred to as *agreement attraction.*

(1) a. The apprentice of the chefs…
b. The apprentice of the chef…

In subsequent research, agreement attraction was also observed in comprehension: It was found that ungrammatical verbs cause little disruption in the same environments where they are erroneously produced. Thus, the ungrammatical *were* in (2a) is read faster than in (2b), as measured in self-paced reading (Lago et al., 2015; Wagers et al., 2009) as well as eye-tracking while reading (Dillon et al., 2013; Jäger et al., 2020; Pearlmutter et al., 1999) studies. In addition, sentences with agreement attraction errors are often perceived, at least momentarily, as grammatical (Franck et al., 2015; Tanner et al., 2014; Wagers et al., 2009), hence the term *grammaticality illusion* (Phillips et al., 2011).

(2) a. The apprentice of the chefs were…
b. The apprentice of the chef were…

Attraction effects have been documented across many languages, interfering with number agreement (Arabic: Tucker et al., 2015, 2021; Armenian: Avetisyan et al., 2020; Dutch: Hartsuiker et al., 2003; English: Bock and Miller, 1991; Wagers et al., 2009; German: Hartsuiker et al., 2003; Hebrew: Deutsch and Dank, 2009, 2011; Hindi: Bhatia and Dillon, 2022; Romanian: Bleotu and Dillon, 2024; Russian: Slioussar, 2018; Spanish: Lago et al., 2015; Turkish: Lago et al., 2019; Türk and Logačev, 2024) and gender agreement (Arabic: Tucker et al., 2015, 2021; French: Vigliocco and Franck, 2001; Hebrew: Deutsch and Dank, 2009, 2011; Italian: Vigliocco and Franck, 1999; Russian: Slioussar and Malko, 2016; Slovak: Badecker and Kuminiak, 2007; Spanish: Antón-Méndez et al., 2002).

These effects have consistent characteristics which suggest that they do not reflect simple lapses of attention, but rather intricate mechanisms of incremental dependency formation. For example, the effects are modulated by the markedness of the distractor's features (Bock and Eberhard, 1993; Wagers et al., 2009; among others) and the morphological overtness of the target's features (Eberhard, 1997; Hartsuiker et al., 2003). Moreover, attraction errors seem to depend on the structural position of the distractor but not its proximity to the verb (Bock and Cutting, 1992; Franck et al., 2002; Vigliocco and Nicol, 1998; Wagers et al., 2009).

Attraction effects in comprehension are usually attributed to memory mechanisms since comprehenders have to accurately maintain and uniquely access a memory representation of the subject to form a subject-verb dependency. Over the years, two main families of accounts have been proposed as explanations for the agreement attraction phenomenon. One type of account attributes attraction effects to the retrieval of the wrong memory item (*retrieval interference accounts*). The other class of accounts attributes attraction effects to difficulties in creating and maintaining a veridical representation of the subject. These accounts, *representational accounts,* hold that interference arises before retrieval and is rooted in how linguistic input is encoded into memory. These families of accounts will be presented briefly in the next two subsections.

1.1.1 Retrieval interference accounts

Retrieval interference accounts of attraction phenomena attribute these effects to a cue-based retrieval mechanism used to form syntactic dependencies. Cue-based retrieval is a general model of the memory mechanisms that support dependency formation. This idea was implemented by Lewis and Vasishth in the ACT-R framework (Lewis and Vasishth, 2005, see also Engelmann et al., 2019; Lewis et al., 2006; Vasishth et al., 2019). In cue-based retrieval models, incoming words are encoded into memory as bundles of structural, morpho-syntactic, and semantic features. Such feature bundles can include, for example, a grammatical number feature (e.g., +/− Plural, in 3), as well as the structural feature indicating the syntactic role of an item (e.g., +/− Subject, but see Arnett and Wagers, 2017). As the input is processed, additional representations are stored in memory, but their activation may decay over time.

(3) The apprentice of the {chefs | chef} were…
[+Subject] [-Subject] [-Subject] *Cue*: +Subject
[-Plural] [+Plural] [-Plural] *Cue*: +Plural

At the point of encountering the verb in the input, the subject must be reactivated to form a dependency with it. The verb thus initiates a search for the subject's memory trace. This retrieval process is guided by a set of retrieval cues associated with the licensing conditions of the verb: the requirement that the target of retrieval occupy a structural subject position [+Subject], that it bears the appropriate agreement features (depending on the inflection of the verb, [+Plural] in 3), and perhaps that it is semantically compatible with the verb (Smith and Vasishth, 2020). Memory items resonate to the cues if their features match them, and this increases their activation. In the Lewis and Vasishth

(2005) model, activated memory items enter a noisy race toward a retrieval threshold, and the most activated item is retrieved, forming a dependency with the verb. The time it takes for the first item to reach the threshold is the retrieval time and is reflected in reading times of the verb (but see Nicenboim and Vasishth, 2018, for a discussion of other implementations of cue-based retrieval mechanisms).

Consider how cue-based retrieval predicts agreement attraction. In (3), the verb *were* has the retrieval cues [+Subject] and [+Plural]. When both nouns are singular, neither matches the agreement cue [+Plural], but *apprentice* matches the structural cue [+Subject]. Thus, the target, i.e., *apprentice*, receives more activation than the distractor, *chef*. In contrast, when the distractor is plural, the target and the distractor both partially match the verb: the target matches the structural cue [+Subject], whereas the distractor matches the agreement cue [+Plural]. Activation in this case would be distributed across both items, such that they both compete for retrieval. The activation levels of the distractor and the target would be set by their baseline activation (before retrieval, affected by a decay function), the weight given to each of the activation cues during retrieval, and trial-to-trial noise fluctuations. When cues distribute activation across both the target and the distractor, this leads to occasional misretrieval of the distractor. In addition, in a noisy race between close competitors, finishing time is faster (on average) than in an unbalanced race, where one competitor is considerably more activated than the other. This is so since, to win a close competition, one competitor has to race particularly quickly ("statistical facilitation"; Raab, 1962; Vasishth et al., 2019). Thus, retrieval times on an ungrammatical verb should be faster when the distractor matches it than when the distractor does not. This can account for agreement attraction effects on reading times.

1.1.2 Representational accounts

In contrast to retrieval-based accounts, representational accounts trace attraction to errors in mapping number and gender features to constituents in memory. This type of interference arises before retrieval is attempted, at encoding. We therefore refer to these as representational distortion accounts (following Hammerly et al., 2019; Yadav et al., 2023). The tradition of representational accounts has multiple proposed mechanisms: stochastic percolation of a feature up the syntactic tree which replaces the original feature value (Bock and Eberhard, 1993; Eberhard, 1997); contribution of plural morphology in some syntactic positions to a scalar value ranging between unambiguously singular and unambiguously plural from which subsequent agreement marking is derived probabilistically (Marking and Morphing: Bock et al., 2001; Eberhard et al., 2005). Those representational accounts were originally framed in the context of agreement production. However, the basic mechanism of misattributing grammatical features to constituents in memory could also lead to the creation of inaccurate memory representations in comprehension.

Recently, we developed a model of representational distortion that focuses on comprehension processes (Keshev et al., 2025). In this model, structure building proceeds as the formation of item-position associations. Namely, to comprehend a sentence, one needs to encode in working memory a transient set of connections that binds every morpheme of the sentence to its syntactic position. The binding of distinct morphemes to similar syntactic positions creates interference. Thus, for example, the plural morpheme of a distractor could contaminate the representation of the number morpheme bound to the target subject (for details about the model and for an account of additional interference effects, see Keshev et al., 2025). Overall, representational accounts are in line with the view that people maintain uncertainty about features of memory items (for review see Bays et al., 2022; Xu and Futrell, 2025) and non-veridical representations of linguistic input in memory (arising due to either fast-and-frugal heuristics: Ferreira and Patson, 2007; or resource-rational processing Futrell et al., 2020).

Stepping back, representational approaches to agreement attraction broadly claim that attraction errors partially reflect uncertainty about which features are associated with items in memory. Thus, if this type of representational uncertainty underlies attraction, then any additional cues (e.g., other agreeing elements) or biases can act to resolve this uncertainty, thereby minimizing the amount of agreement attraction. We return to the idea of uncertainty and utilization of regularities to mitigate distortion and attraction in Section 1.2.3.

1.2 Previous evidence for representational accounts

While there is widespread support for retrieval-based approaches to attraction, recent findings offer renewed interest in representational distortion models and provide evidence for some unique predictions of representational distortion accounts in agreement attraction data. Such evidence comes from the final interpretation of attraction sentences and the effects of multiple agreement markings throughout the sentence. The following subsections discuss the above-mentioned types of evidence, as well as proposals, in the context of the grammaticality asymmetry, that the retrieval and distortion accounts are not mutually exclusive.

1.2.1 The grammaticality asymmetry

The question of whether agreement errors reflect retrieval errors, representation errors, or both has been strongly influenced by the observation of a *grammaticality asymmetry* in agreement attraction effects in comprehension. The grammaticality asymmetry refers to the fact that while sentences such as (4) are readily perceived as grammatical (i.e., exhibit a grammaticality illusion), it is somewhat less common to see a mirror image 'ungrammaticality illusion' in grammatical sentences in which a distractor noun does not match the verb, as in (5) (Wagers et al., 2009).

(4) The apprentice of the chefs were experienced.

(5) The apprentice of the chefs was experienced.

This apparent lack of an ungrammaticality illusion constitutes a major challenge for representational theories of attraction (see e.g., Wagers et al., 2009). For example, according to Marking and

Morphing, the scalar number of the NP in (4)–(5) is somewhere between singular and plural, leading to a certain probability of production of a plural verb as in (4). Accordingly, in precisely this proportion of cases, (5) should be perceived as ungrammatical, due to a number mismatch between the subject and the verb number. This ungrammaticality illusion, however, was not observed in Wagers et al. (2009), as well as in other studies (e.g., Lago et al., 2015; Tucker et al., 2015).

However, the question of whether illusions of ungrammaticality occur in grammatical sentences has been reopened in recent literature. Recent work on this topic suggests that the ungrammaticality illusion is detectable and that task or response artifacts contribute to its elusiveness. Hammerly et al. (2019) argued that in speeded acceptability judgments, the grammaticality asymmetry may reflect response bias. They exhibited in modeling work that when participants are biased to respond 'grammatical', a grammaticality asymmetry arises even if the underlying perception of well-formedness is affected in grammatical sentences as well. Moreover, in a series of acceptability judgment experiments, Hammerly et al. showed that as positive response bias was neutralized, an illusion of ungrammaticality was observed. Laurinavichyute and von der Malsburg (2024) provided evidence that the illusion of ungrammaticality in reading time measures may also be task dependent: In experimental contexts where participants expected to judge a sentence's grammaticality, they slowed down when reading a verb that mismatched a distractor in number (analogous to 5), but this effect went away when comprehenders were not expecting to engage in grammaticality judgements. Like Hammerly et al. (2019), this finding suggests that the illusion of ungrammaticality predicted by representational accounts is attested, even if it is only observed in some experimental contexts. These studies thus suggest that representational distortion contributes to agreement attraction effects, possibly in addition to retrieval interference.

Finally, Yadav et al. (2023) proposed a hybrid model of agreement attraction based on the elusiveness of the ungrammaticality illusion. Yadav and his colleagues proposed that both representational distortion and retrieval interference conspire to produce the illusion of grammaticality (i.e., in ungrammatical sentences). At the same time, these forces act in opposite directions in grammatical sentences, thus making the ungrammaticality illusion untraceable. Yadav et al. (2023) examined the fit of different computational retrieval-based, representational, and hybrid models of attraction. They found that the RT data from a series of reading time datasets testing for attraction in grammatical and ungrammatical sentences were best captured by models that incorporated both representational distortion and retrieval interference (but see Laurinavichyute and von der Malsburg, 2024). All in all, a range of recent evidence suggests that representational distortion contributes to agreement attraction as well.

1.2.2 Comprehension errors

Another key finding that provides evidence for representational distortion effects comes from the patterns of comprehension errors in attraction configurations. Most studies on agreement attraction in comprehension examine reading latencies or accuracy in acceptability judgements. However, a crucial question that arises from the attested illusion of grammaticality is how these illusory sentences are interpreted during and after reading.

Representational accounts make the strong prediction that a plural distractor may cause a singular head noun to be misinterpreted as plural. If agreement attraction arises when the head noun is inaccurately encoded as plural, then we should find evidence that readers recall a non-veridical representation of the subject as plural, i.e., remember the subject in "the apprentice of the chefs were…" as 'apprentices'. Cue-based retrieval, in contrast, predicts that interpretation will involve only veridical representations, since this model assumes accurate and fixed memory encoding in current implementations. In illusory sentences, cue-based retrieval predicts increased rates of falsely interpreting the distractor as the verb's subject, as the distractor was retrieved at the verb. Thus, in "the apprentice of the chefs were…", 'chefs' will be interpreted as the subject.

Several studies have examined final interpretation errors to test the predictions of representational and retrieval-based approaches. Patson and Husband (2016) presented participants with questions like "Was there more than one key?", following sentences such as "The key to the cabinet/s was/were lost." They found that readers incorrectly answered affirmatively more when a plural distractor was present (see also Brehm et al., 2019). Brehm et al. (2021) examined misinterpretations in a visual world task involving an array of plural and singular depictions of the subject and the distractor. They found that misinterpretations of a head noun's number occur online: fixations to "keys" increased following a plural distractor ("the key to the cabinets…"). A sentence-final forced-choice task similarly revealed that plural distractors increased the odds of misinterpreting the subject as plural. Lastly, Paape et al. (2021) examined the interpretation of agreement attraction sentences in Eastern Armenian using open-ended questions. The authors coded responses for errors in the subject's identity and number. They found that the distractor's features affected the rate of number errors, and that the verb's features did not affect misremembering the distractor as the subject ("cabinets" instead of "key"). The finding that a plural distractor noun may cause a singular head noun to be wrongly interpreted as plural, both during and after reading, supports representational accounts and is not predicted by cue-based retrieval. Thus, this finding provides evidence for the unique contribution of representational distortion to attraction effects.

Interestingly, a recurring finding across these studies is that number misinterpretations of the subject head also depend on subject-verb mismatches, in addition to subject-distractor mismatches. Ungrammatical number marking on the verb leads to increased error rates in a number-judgment task (Brehm et al., 2019; Patson and Husband, 2016), looks to a number competitor in visual world tasks (Brehm et al., 2021), rates of non-veridical subject responses in four-alternative forced choice tasks (Brehm et al., 2021), and in free recall tasks (Paape et al., 2021). This raises the possibility that the representation of the subject's number features can be modulated both while it is being maintained in working memory and during the formation of the dependency with the verb. That is, additional evidence that the subject is plural, in the form of morphological agreement on the verb, appears to meaningfully change the likelihood of construing the subject as plural (see also Molinaro et al., 2008). As we explain in the next subsection, findings that memory representations may be dynamically updated as input unfolds provide further support for representational interference.

1.2.3 Updating feature representations during dependency formation

A major disparity between representational approaches and pure retrieval-based approaches has to do with the stability of memory representations. In cue-based retrieval, the feature bundles that constitute memory items are fixed. In contrast, in representational approaches, associations between features and items in memory can shift, and comprehenders may maintain uncertainty about their memory representations. This does not only mean that memory contents can be distorted but also that linguistic regularities, namely grammatical feature markings, can facilitate veridical encodings recovering memory items (Brady et al., 2009; Norris and Kalm, 2021). Here, we review findings that verbal agreement marking can induce shifts in the representation of the subject's features. These include recovery of veridical representations despite the distractor's effect, on the one hand (given a grammatical verb), and distortion effects irrespective of the presence of a distractor, on the other hand (given an ungrammatical verb). Such findings can be taken to support representational approaches as they suggest that (i) feature-item bindings in memory are not always fixed; and that (ii) uncertainty is maintained and controlled based on linguistic regularities. Both are in line with representational distortion models as discussed in Section 1.1.2.

The first finding indicating that verbal agreement marking can modulate the representation of the subject comes from research on reflexive processing. Molinaro et al. (2008) investigated the processing of reflexive pronouns following a subject-verb mismatch. They found that reflexives mismatching the verb but matching the subject (6a) were perceived as ungrammatical (as indicated by a P600 ERP effect), whereas reflexives matching the verb but mismatching the subject were experienced as grammatical (6b). The authors therefore suggest that readers coerce the representation of the subject to match features of the verb in real time and that this affects predictions as to the reflexive form. The preference for reflexive-verb match over reflexive-subject match was also replicated in reading times (Keshev and Meltzer-Asscher, 2024). These findings are not surprising under representational distortion approaches, since such approaches assume that local material (here, the verb) may alter the representation of a previously encountered element (here, the subject)—resulting in a reflexive-verb number match being more important than a reflexive-subject mismatch.

(6) a. The famous dancer were nervously preparing **herself** to face the crowd.
b. The famous dancer were nervously preparing **themselves** to face the crowd.

Verbal agreement marking was also found to mitigate vulnerability to attraction in subsequent agreeing sites. Keshev and Meltzer-Asscher (2024) found that in sentences with overt verbal agreement marking, attraction at an ungrammatical reflexive pronoun was reduced compared to sentences without verbal agreement marking. This finding supports the hypothesis that readers use agreement cues on the verb to update the representation of its subject. Specifically, on this view, the update contributed by an inflected verb's features overrides a potentially inaccurate representation of the subject noun's features, thus sharpening the reader's confidence in the number representation of the subject and subsequently making readers less susceptible to attraction from distractor nouns. This result too is unexpected under cue-based retrieval (see explanation in Keshev and Meltzer-Asscher, 2024, and in the next subsection).

1.3 The current study

In the current study, we aim to expand the agreement updating picture and provide further tests of the predictions made by representational distortion approaches to attraction. As explained above, Keshev and Meltzer-Asscher (2024) have provided evidence that comprehenders use agreement marking at the verb to update memory representations. However, they found this in the context of reflexive attraction. As a reflexive object is part of the verb phrase, agreement updating could be limited to cases where there is a direct grammatical relation between the retrieval site and the preceding agreement-marking/updating site, as arises between an object reflexive and its verb. In the current study, we test whether the effect of earlier agreement marking on subsequent attraction extends to other configurations. We use VP coordination to create two agreement dependencies with the same subject noun (7a).

(7) a. The apprentice of the {chef | chefs} <u>works</u> diligently and **were recruited** by a top restaurant. WithCue
b. The apprentice of the {chef | chefs} <u>worked</u> diligently and **were recruited** by a top restaurant. NoCue

In this configuration, the verb of the first conjunct (underlined) presents grammatical agreement with the subject. Would the second verb (in bold) in such a sentence be vulnerable to attraction? Under representational approaches, the distractor noun 'chefs' may disrupt the memory representation of the subject 'apprentice', causing it to bear the wrong number feature. In such a case, the comprehender could use the verbal agreement marking in the first conjunct to amend this memory disruption. Such feature updating would decrease the vulnerability of the second verb to representational attraction. This predicts lower rates of attraction at the second verb in sentences where the first verb carries grammatical agreement marking (7a), relative to cases where the first verb does not carry overt agreement marking (7b).

On the other hand, in a pure retrieval-based model where memory representations are fixed, there is no uncertainty about the features of the subject. In such a model, the only possible errors are misretrieval errors—access to the distractor instead of the subject. In Cue-Based Retrieval, one dependency can affect the rates of subsequent attraction in another dependency only indirectly via modulating the activation levels of memory items. Reactivation of the subject at the first verb can somewhat decrease attraction as it raises the activation of the subject, making it a stronger competitor at retrieval. The extent to which attraction is blocked following recent activation depends on decay and noise parameters

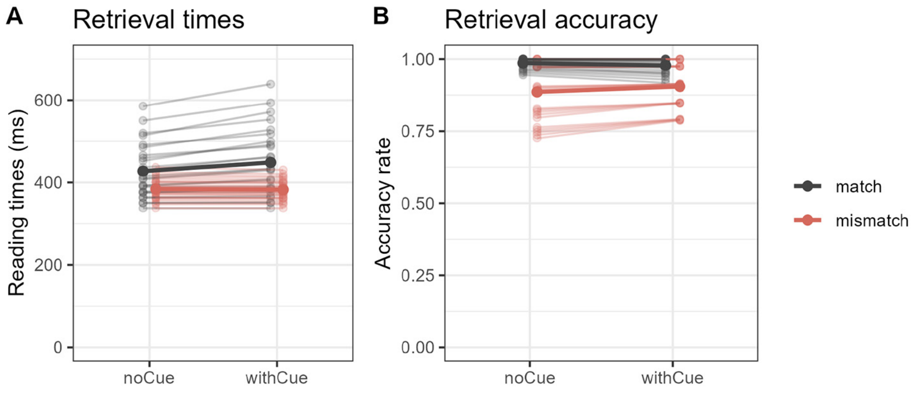

FIGURE 1
Results of cue-based retrieval simulations for reading times **(A)** and accuracy measures **(B)**. The withCue conditions correspond to sentences like (7a) and noCue to (7b). Match/mismatch refers to the feature configuration of the distractor relative to the target, e.g., in example (7), the singular *chef* would reflect feature match with the target (*apprentice*), while the plural *chefs* would constitute a mismatch. Opaque dots reflect individual simulations across different values of the decay parameter (between 0.5 and 0.9) and of the noise parameter (between 0.1 and 0.5). Lines connect simulations of the same parameter values. The solid dots reflect the grand mean. The attraction effect is reflected in the distance between the black and red lines. Parallelism of these lines suggests that attraction effects are consistent for sentences with and without agreement marking at the first conjunct.

(see Figure 1 and Dillon, 2011). However, verbs carrying agreement marking (7a) and those that do not (7b) would differ very little in terms of blocking later attraction, as both consistently activate only the subject. This predicts similar attraction rates across conditions. These predictions for attraction at the second conjunct are portrayed in Figure 1, derived from simulations of the cue-based retrieval model.

Across a series of studies, we find some evidence for, and some failures to find evidence for feature updating. On balance, we think feature updating is a promising theory in light of our results, and deserving of further research. In addition, we find exploratory evidence for an illusion of ungrammaticality and for number misrepresentation in final comprehension. We argue that some form of feature distortion has to be assumed to fully capture these attraction phenomena.

1.3.1 Data availability

Data, materials, and analysis code associated with this study are available through OSF at https://osf.io/gdjne/. Experiments 1 and 2 were pre-registered. The preregistrations are available at https://osf.io/njubx and https://osf.io/9x7gp, respectively.

2 Experiment 1: evidence for feature updating in forced verb form choice

2.1 Methods

2.1.1 Participants

We recruited participants until reaching our preregistered cap of 60 participants who passed the exclusion criteria. We recruited a total of 88 self-reported native English-speaking participants through the Prolific Academic online platform. Participants gave informed consent and received monetary compensation of 3.34 USD for their participation (a rate of approximately 13 USD/h).

TABLE 1 Example of an experimental item set from Experiment 1.

Distractor-subject features	Sentence	Alternative completions
Match	The apprentice of the chef {worked | works} diligently an	was were
Mismatch	The apprentice of the chefs {worked | works} diligently and	

The intermediate verb, where the availability of verbal agreement marking was manipulated, is underlined. The bracketed verbs represent no-cue and with-cue conditions correspondingly.

This experiment was determined to be exempt research by the Institutional Review Board of the University of Massachusetts.

2.1.2 Materials

We constructed 36 item sets of four conditions. Items followed the general structure in (7). In a 2 × 2 design, we manipulated the number feature of the distractor noun (matching or mismatching the subject) and the availability of agreement cues in the first conjunct (with or without a cue). Overt marking of verbal agreement in the first conjunct was manipulated using the tense of the verb (present or past tense, see Table 1). Participants were asked to choose between a singular and a plural verb form as the next word of the sentence. In half of the experimental items (18 sets), the choice was between forms of an auxiliary verb (*was* vs. *were*), and in the other half (18 sets), the choice was between forms of an intransitive lexical verb (e.g., *smiles* vs. *smile*).

The experimental items were distributed across four Latin Square lists. Each list was combined with the same set of 114 filler items. The filler items aimed at balancing the correct agreement choices and preventing strategic task behavior, such as always selecting the singular form or always agreeing with the first noun of the sentence. Six simple filler items were designated catch trials.

Catch trials were simple mono-clausal sentences with no distractor nouns (e.g., "The old and wise president [worries, worry]").

2.1.3 Procedure

The experiment included rapid serial presentation of a preamble and a forced completion task. Sentence preambles were presented one word at a time, for a duration of 250 ms per word with an inter-stimulus interval of 150 ms. Participants were then prompted to select from two verb forms presented on the screen, a singular or plural verb. The trial was terminated if no response was made within 3 s. No feedback on response accuracy was provided.

The experiment was implemented in PCIbex (PennController for Internet-Based Experiments, Zehr and Schwarz, 2018). Participants performed the experiment remotely on their own computer. Before starting the experiment, participants undertook a practice block of five sentences. The experiment took approximately 15 min to complete.

2.1.4 Data analysis

Participants were excluded from the analysis if they failed to provide a coherent English response to a preregistered open-ended prompt (one participant) or failed more than one of the catch trials (26 participants). For the remaining 61 participants, we excluded from analysis data points with response times below 100 ms (affecting 0.18% of the data).

The data were analyzed in R (R Development Core Team, 2015) using Bayesian hierarchical models with a Bernoulli link function. We fitted Bayesian hierarchical models in Stan (Carpenter et al., 2017) via the brms package (Bürkner, 2017). We used sum coding for both experimental factors (½ for match and for with-cue; –½ for mismatch and for no-cue conditions).[1] In addition, to evaluate the size of attraction effects, we fitted another model with nested contrasts. This model included the main effect of cue availability and two pairwise comparisons, reflecting the attraction contrast (match vs. mismatch) within each level of the cue availability factor. Both models included the maximal random effects structure by-items and by-participants, including random intercepts and random slopes for all fixed effect predictors (main effects and interactions). We report the posteriors' 95% credible interval (CrI) and take it to support the presence of an effect if it excludes zero.

We use weakly informative priors: a standard normal distribution, $N(0, 1)$, as the prior for fixed effects and for the standard deviation parameters; a normal prior of N(0,3) for the intercept; and the LKJ prior for correlation matrices of random effects (Lewandowski et al., 2009). Four Monte Carlo Markov Chains of 4,000 iterations each were sampled from the posterior distribution. The first 2,000 samples of each chain were discarded as a warm-up. Convergence was checked using the R-hat statistic, which was at 1.0 for all fixed effects.

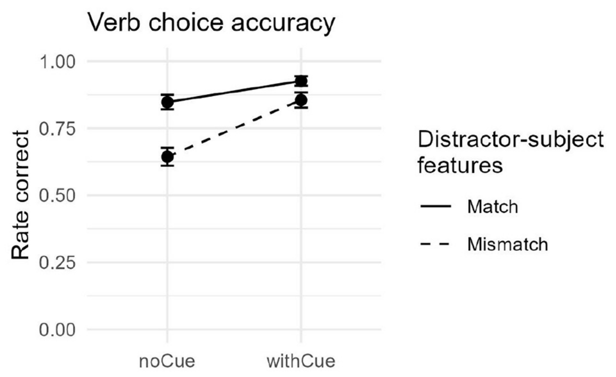

FIGURE 2
Mean accuracy in Experiment 1 by condition. Rates of the correct (singular) verb in the two-alternative forced-choice task. Error bars represent +/−SE.

TABLE 2 Results of the analysis of Experiment 1.

Main effect model		Nested contrasts model	
Cue availability	1.39 [0.90, 1.90]	Cue availability	1.37 [0.86, 1.9]
Attraction	1.33 [0.94, 1.75]	Attraction without a cue	1.63 [1.11, 2.19]
Interaction	−0.63 [−1.4, 0.15]	Attraction with a cue	0.89 [0.26, 1.50]

Mean and 95% credible interval (on the log-odd scale) of the posterior distribution for fixed effects (under the weakly informative priors set).

We also calculate Bayes Factors (BF) to evaluate the evidence for the critical interaction. BFs were computed using the bridgesampling R package (Gronau et al., 2020). For a stable calculation of BF, the number of iterations was increased to 10,000 (of which 2,000 were warm-up). We follow the common guidelines for interpretation of BFs (Lee and Wagenmakers, 2014), whereby BFs above 3 or below 0.33 are considered moderate evidence, and BFs above 10 or below 0.1 are considered moderate strong evidence.

2.2 Results

Accuracy rates in the different conditions are presented in Figure 2. Model results are summarized in Table 2.

The credible interval and the Bayes factor analyses revealed conflicting results concerning the interaction between the number of the distractor and the availability of an agreement cue on the first verb. While the credible interval of this posterior crossed zero, the Bayes Factor analysis provided strong evidence for an interaction effect. This evidence was observed both under weakly informative priors (BF = 18, in favor of an effect) and informative ones (BF = 39.9, in favor of an effect). This interaction reflects a more prominent attraction when no agreement cue is available in the first conjunct compared to when it is available. The nested contrasts model supported the observation that a mismatching distractor impaired accuracy in no-cue conditions (mean posterior of the attraction pairwise contrast: 1.63; CrI: [1.11, 2.19]) somewhat

[1] We diverge from the pre-registered contrast coding in that we use ½ and -½ rather than 1 and −1 for sum coding (or 1/3 and −2/3 instead of 1 and −2 for Helmert coding in later experiments). This was done to make sure that the estimates of the model reflect the effect of interest rather than half the effect size, and has no effect on the position of the posterior distribution relative to zero.

more than in with-cue conditions (posterior mean [CrI]: 0.89 [0.26, 1.50]).

2.3 Discussion

Experiment 1 provides some limited support for the claim that agreement cues can be used to reduce the uncertainty about number feature-item bindings in memory. Based on our Bayes Factor analysis, the presence of overt agreement marking on an intermediate verb reduces vulnerability to attraction at a subsequent verb. This is compatible with a model where features can be distorted in memory. In models of representational distortion (e.g., Marking and Morphing, feature percolation), additional cues for the features of the subject can reduce uncertainty about its memory representation. Such a process would (partly) block the attraction effects of the representational distortion kind. On the other hand, the findings conflict with models like Cue-Based Retrieval. Under Cue-Based Retrieval, representations are fixed and therefore cannot be updated. In addition, updating of feature representations should not prevent attraction in this model, as it derives attractions from erroneous retrieval of the distractor rather than from distortion of the subject.

However, three features of our results hinder strong conclusions. First, the criteria of a credible interval excluding zero did not support an interaction effect. The was only 94.2% chance of an effect [$\Pr(\beta < 0) = 0.942$]. Second, as explained in Section 1.3, these results can be thought to still be compatible with the cue-based retrieval model if specific parameter settings are invoked. To further tease apart cue-based retrieval and representational accounts of attraction, in the following experiments (Experiment 2–3), we examine whether feature updating is detected in incremental dependency formation in reading times. Lastly, it should be mentioned that a Hebrew version of Experiment 1 was pre-registered and run within this project. We deem the Hebrew experiment inconclusive, as accuracy rates in this experiment were very high across conditions, rendering attraction effects rather modest and giving rise to large CrIs (additional information is available on our OSF repository)[2].

3 Experiment 2: exploratory evidence for an ungrammaticality illusion in eye-tracking while reading

In Experiment 2, we probe the effects of feature updating on attraction using reading time measures, where predictions of the cue-based retrieval model are more consistent (see Figure 1). This experiment uses the same coordinate structure introduced in Section 1.3: the first conjunct hosts a grammatical verb which could carry over agreement marking (e.g., the present tense form *works*) or not (e.g., the past tense form *worked*). Attraction effects, as reflected in facilitatory interference on an ungrammatical verb, are measured in the second conjunct.

TABLE 3 Example of an experimental item set from Experiments 2–3.

Agreement condition	Sentence
Grammatical baseline	The apprentices of the chef {worked | work} diligently and **were recruited** by a top restaurant.
Ungrammatical mismatch	The apprentice of the chefs {worked | works} diligently and **were recruited** by a top restaurant.
Ungrammatical match	The apprentice of the chef {worked | works} diligently and **were recruited** by a top restaurant.

The critical region is in bold. The intermediate verb that could license verbal agreement cues is underlined.

If attraction arises from competition between the target and the distractor as part of cue-based retrieval, we should observe at the second verb similar attraction effects, whether or not overt verbal agreement was marked in the first conjunct (see Figure 1 for simulations). In contrast, if attraction arises from representational distortion, overt agreement marking on the first verb could allow the parser to update its representation of the subject (correcting its feature array). In that case, overt agreement marking in the first conjunct would reduce the vulnerability to attraction at the second conjunct.

3.1.1 Participants

We recruited participants until reaching our preregistered cap of 96 participants who passed the exclusion criteria. We recruited a total of 118 self-reported native English-speaking participants from the student body of the University of Massachusetts Amherst. Participants gave informed consent and received either course credit or monetary compensation of 15$ for their participation. This experiment was approved by the Institutional Review Board of the University of Massachusetts.

3.1.2 Materials

Items were based on those of Experiment 1 (see Table 3). We added 4 sets and adapted the resulting 36 sets to comply with a 3 × 2 design. In all items, the critical verb (in the second conjunct) was a plural auxiliary (were). We manipulated the number features of the subject and the distractor to vary the grammaticality of the second (plural) verb (plural vs. singular subject) and attraction (feature match between the distractor and the subject). Overall, the agreement manipulation included three levels: a grammatical sentence (i.e., with a plural subject head); an ungrammatical mismatch sentence, with the attraction-prone configuration of a singular subject head and a plural distractor; and an ungrammatical match sentence, with a singular subject and a singular distractor. As in Experiment 1, the design manipulated the availability of agreement cues in the first conjunct using past vs. present tense verbs to evaluate feature updating.

A comprehension question followed each sentence. Comprehension questions took the form of a four-alternative forced choice (4AFC) task. Questions about the experimental items

[2] For example, the interaction's CrI had a range of 2.35 (on the log odd scale) in the Hebrew experiment. In contrast, in the English version this range was only 1.55. Thus this experiment failed to detect evidence for or against the interaction (BF of 1.25 for an effect was with weakly informative priors and 2.77 with informative priors).

targeted the subject of the second verb, as in (8). Participants were asked to select from four alternatives, which included plural and singular versions of the subject and the distractor.

(8) Who was recruited by a top restaurant?

The apprentice	The chef
The apprentices	The chefs

The experimental items were distributed across six Latin Square lists. Each list was combined with the same set of 124 filler items. Of those fillers, 20 were designated catch trials. Catch trails were either short mono-clausal sentences with non-reversible roles (e.g., "The excited girl bought the shiny toy", with the question "Who bought the toy?", the girl/the girls/the boy/the boys), or longer sentences designed to resemble experimental items but with questions probing the distractor rather than the subject (e.g., "The brothers of the volunteers came over and asked them to return home.", with the question "Who was asked to return?", the brother/the brothers/the volunteer/the volunteers).

3.1.3 Procedure

The experiment was an eye-tracking while reading experiment. Eye movements were monitored with a tower-mounted EyeLink1000. The experiment was implemented in SR Research's Experiment Builder. The participants were seated at a distance of 60 cm from a presentation monitor (size: 432 × 216 mm; resolution: 1,600 × 900). Head movements were restricted using a chinrest. The sentences were presented on the screen in a monospaced font of size 14, resulting in 3.5 characters within one degree of visual angle. Before starting the experiment, participants undertook a practice block of five sentences. A break was offered after a third and after two-thirds of the experiment. Participants were invited to take additional breaks whenever needed. After each break, recalibration was performed. The experiment session lasted approximately 1 h.

3.1.3 Data analysis

We excluded from the analysis trials with a first pass blink or track loss on the critical region. Participants were excluded from the analysis if more than 25% of their data was lost due to track loss or blinks (8 participants) or if they failed more than 6 (30%) of the catch trials (14 additional participants). We analyze the main eyetracking measures: first pass reading times, go past reading times (regression path), proportion of first-pass regressions out, and total reading times. Results of the 4AFC task are analyzed separately in Section 5.

We fitted Bayesian hierarchical models with a lognormal link function for reading time measures and models with a Bernoulli link function for regression proportions. We used sum coding for the cue availability factor (½ for with-cue and –½ for no-cue). For the three-level agreement factor, we used Helmert coding (Schad et al., 2020) that produced two predictors: Grammaticality, contrasting the grammatical baseline condition (–2/3) with the mean of the ungrammatical conditions (1/3 each); and attraction, contrasting the two ungrammatical conditions (match coded as ½ and mismatch as –½). To examine effects within the no-cue and within the with-cue levels, we also implemented a nested contrasts model. In this model, we included a main effect of cue, pairwise comparisons of attraction (mismatch vs. match) within each of the cue availability levels, and simple effects of grammaticality within each of the cue availability levels. All models included the maximal random effect structure by-items and by-participants, including random intercepts and random slopes for all fixed-effect predictors. Modeling parameters were the same as in Experiment 1, except for the prior of the intercept used for reading time measures (first pass, go past, and total reading times), which was wider: $N(0, 10)$. In addition, summary statistics of these measures' posteriors were transformed back to the millisecond (ms) scale for ease of interpretation.

3.3 Results

3.3.1 Pre-registered analysis: Reading times at the second verb

Our preregistered analysis concerned the region of the second verb phrase, including the number-marked auxiliary ('were') and the past participle following it (e.g., 'recruited'). We analyze the different eye tracking measures in a main effect model and a nested comparisons model (see Table 4). Total reading times, where attraction often occurs, are presented in Figure 3 (across regions) and Figure 4 (a close-up of the critical region), and the results of the statistical analyses are presented in Table 4.

Grammaticality effects were detected across all measures, with faster reading and fewer regressions for grammatical compared to ungrammatical verbs. In contrast, we failed to detect an attraction effect (i.e., reduced reading times in ungrammatical mismatch compared to match conditions) in any of the measures. We also failed to observe the predicted interaction with cue availability. The CrI of this effect crossed zero across all measures.

To quantify the evidence for/against a Cue × Attraction interaction, we calculated Bayes Factors (BFs) for total times, where attraction has being detected in previous studies (Jäger et al., 2020). BFs were computed using the bridgesampling R package (Gronau et al., 2020). We follow the common guidelines for interpretation of BFs (Lee and Wagenmakers, 2014), and consider a ratio between 3 and 10 as moderate, between 10 and 100 as strong, and above 100 as extreme. As BFs are sensitive to the prior distribution (Gelman et al., 2017), we computed BFs for a range of plausible priors: the weakly informative priors included in the basic analysis—$N(0, 1)$ for all fixed effects, as well as a more informative prior of $N(0, 0.5)$ and $N(0, 0.1)$ on the critical interaction (priors associated with other fixed effects and with the random effects were kept identical).

The Bayes Factor analysis produces strong to extreme evidence for the null. The ratio between the data's likelihood under the null model (excluding the critical interaction) and under the model that includes all main effects and interactions (the hypothesis model) was: 12.62, under the most informative prior; 444 under the intermediate prior, and 17,212 under the weakest prior.

Visual examination of the results suggests unexpected reading patterns at the first verb and at the subject phrase. We therefore statistically analyze these regions as well. Details regarding the exploratory analysis of the subject region are available in Appendix A. The next subsection focuses on the exploratory analysis of the verb region.

TABLE 4 Results of the pre-registered analysis of Experiment 2.

Contrast label	First pass	Regressions out	Go past	Total time
M1: Cue availability	8 [−5, 21]	0.03 [−0.22, 0.27]	19 [−7, 44]	8 [−17, 34]
M1: Grammaticality	17 [3, 31]	0.38 [0.09, 0.68]	46 [21, 70]	105 [77, 133]
M1: Attraction	14 [−1, 30]	0.11 [−0.16, 0.37]	22 [−4, 47]	29 [−6, 64]
M1: Interaction, Cue × Gram	20 [−7, 46]	−0.03 [−0.54, 0.49]	20 [−20, 69]	18 [−52, 87]
M1: Interaction, Cue × Attract	12 [−20, 43]	−0.01 [−0.55, 0.52]	18 [−52, 87]	15 [−56, 86]
M2: Cue availability	8 [−5, 20]	0.00 [−0.27, 0.26]	19 [−7, 44]	8 [−18, 34]
M2: No cue grammaticality	6 [−13, 25]	0.38 [0.01, 0.76]	36 [6, 67]	95 [59, 131]
M2: With cue grammaticality	30 [11, 49]	0.42 [−0.04, 0.92]	60 [25, 94]	115 [78, 151]
M2: No cue attraction	8 [−14, 31]	0.06 [−0.36, 0.47]	16 [−20, 52]	20 [−26, 60]
M2: With cue attraction	18 [−5, 42]	0.04 [−0.37, 0.42]	17 [−24, 57]	38 [−15, 90]

Mean and 95% credible interval of the posterior distribution for fixed effects (on the ms scale for reading times and on the log-odd scale for regression proportion). Credible intervals that do not cross zero are shaded gray.

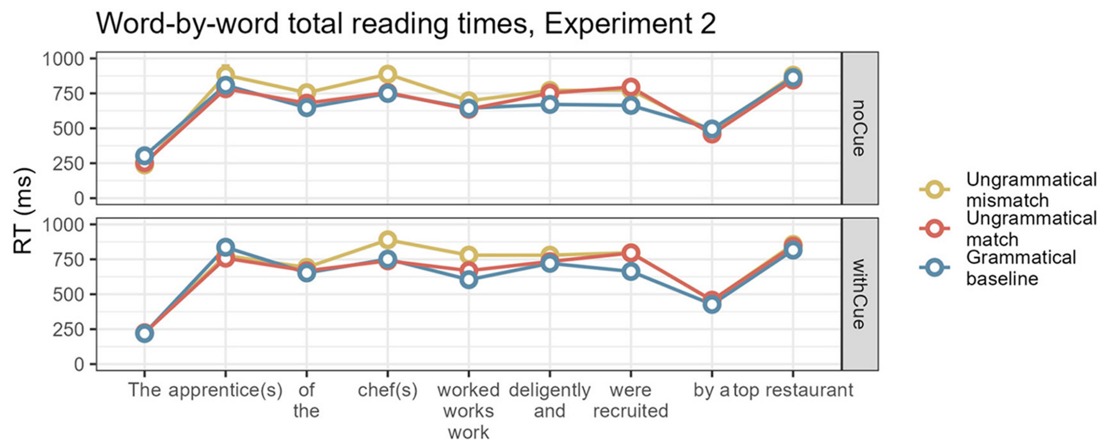

FIGURE 3
Word-by-word total reading time by condition, Experiment 2.

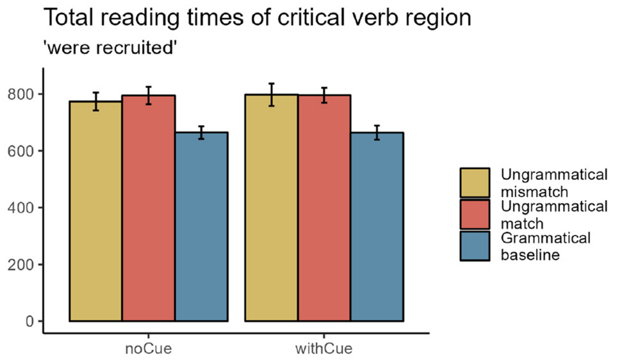

FIGURE 4
Mean total reading times by condition, at the second verb region. Error bars represent +/−SE.

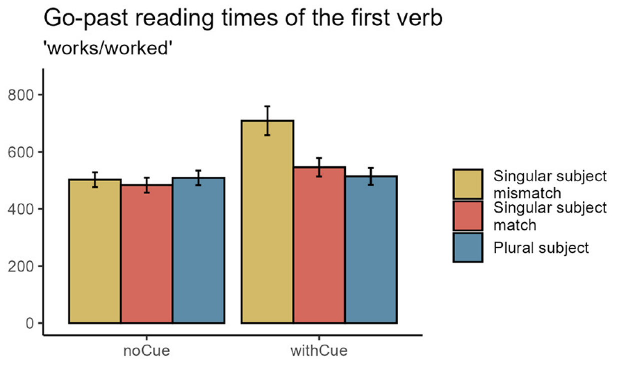

FIGURE 5
Mean go-past times (regression path duration) by condition, at the first verb. Error bars represent +/−SE.

3.3.2 Exploratory analysis: reading times at the first verb

We examined early measures of reading (first fixations, regressions out, and go-past times) at the verb of the first conjunct to evaluate how experimental stimuli were processed before the critical region (see Figure 5 and Table 5). We focus on measures that trace readers' behavior before they exit the region to the right to avoid contamination from the processing of the second

TABLE 5 Results of an exploratory analysis of the first verb region, Experiment 2.

Contrast label	First pass	Regressions out	Go past
M1: Cue availability	−4 [−16, 7]	0.38 [0.15, 0.63]	39 [12, 66]
M1: Plurality	8 [−4, 20]	0.02 [−0.20, 0.24]	22 [−1, 34]
M1: Attraction	−11 [−26, 4]	−0.21 [−0.47, 0.03]	−46 [−72, −20]
M1 Interaction: Cue × Plural	20 [−2, 43]	0.36 [−0.13, 0.24]	81 [33, 128]
M1 Interaction: Cue × Attract	−19 [−45, 8]	−0.27 [−0.79, 0.24]	−65 [−122, −8]
M2: Cue availability	−4 [−16, 8]	0.37 [0.13, 0,61]	40 [13, 66]
M2: No cue plurality	−2 [−19, 15]	−0.19 [−0.50, 0.12]	−19 [−49, 11]
M2: With cue plurality	18 [1, 35]	0.20 [−0.14, 0.55]	62 [28, 97]
M2: No cue attraction	−2 [−21, 17]	−0.07 [−0.44, 0.29]	−13 [−47, 21]
M2: With cue attraction	−21 [−41, −1]	−0.35 [−0.73, 0.01]	−79 [−126, −34]

Mean and 95% credible interval of the posterior distribution for fixed effects (on the ms scale for reading times and on the log-odd scale for regression proportion). Credible intervals that do not cross zero are shaded gray. M1 is the main effects model. M2 is the nested contrasts model.

conjunct. Note that at the point of the first verb, all conditions are grammatical—this verb either matched the subject in number or carried no agreement marking (in no-cue conditions). The main difference between "grammatical" and "ungrammatical" conditions at this point is whether the subject was singular or plural. Conditions that later resolved as completely grammatical included a plural subject, while conditions where the second verb is ungrammatical had a singular subject. Therefore, and to avoid confusion, we label the contrast between the grammatical condition and ungrammatical conditions "effect of subject plurality".

We observed a main effect of attraction in go-past times, such that the verb was read more slowly when the subject was singular and the distractor was plural. This main effect was qualified by an interaction with cue availability: the increase in reading times following mismatching distractors arose only in the with-cue condition, namely with an inflected verb. The nested model for this measure revealed that the inhibitory attraction effect occurred only when the first verb carried agreement features. Yet it should be noted that the Bayes Factor for the interaction of attraction and cue availability in this region was not conclusive. Across three increasingly informative prior sets, the BF was in the range of anecdotal evidence for the null (all three BFs were between 1 and 2.5 for the null).

If this pattern slowdown at the first verb reflects a reliable effect, it suggests a penalty related to the mismatch between the distractor and the grammatically inflected verb: Only in with-cue conditions did the verb in this region carry agreement features. The fact that the slowdown occurs in this condition, but not when the verb lacks overt agreement marking, suggests that the penalty is not associated with the features of the distractor *per se*, such as a delayed plural complexity effect (see Wagers et al., 2009). The fact that this effect is selective to inflected verbs suggests that it reflects something about the relationship between the distractor and the verb. Therefore, we suggest that this finding reflects an ungrammaticality illusion whereby grammatical verbs incur a penalty for mismatching a distractor—an effect which is theoretically predicted by representational accounts but is not commonly observed (see Section 1.2.1).

The analysis in Table 5 also indicated a main effect of cue availability in regressions out and go-past times, such that past tense verbs were read faster than agreeing verbs. The nested contrasts model also revealed an effect of plurality within the with-cue conditions in first pass and go-past times, such that the reading time of a plural verb was faster than that of singular verbs (across match and mismatch conditions). The effects of cue availability and the nested plurality effect seem to be driven by a slowdown in the mismatch condition of with-cue sentences (see Figure 6). Thus, they are likely to reflect the same ungrammaticality illusion discussed above.

3.4 Discussion

In Experiment 2, we failed to detect attraction in conditions with and without intermediate agreement marking. Due to our failure to detect the basic agreement attraction effect, we take these results to be inconclusive as to the question of agreement updating during incremental dependency formation. In Experiment 3, we re-examine whether agreement updating (as observed in Experiment 1) can be observed in reading times and during real-time dependency formation (using self-paced reading).

Despite the lack of evidence for agreement updating, Experiment 2 did produce exploratory evidence for representational accounts of attraction. We found some evidence for an ungrammaticality illusion at the first verb (in line with Laurinavichyute and von der Malsburg, 2024). In our data, grammatical verbs were read more slowly when they mismatched features of the distractor. This effect was detected as an interaction in the credible interval analysis, but the Bayes Factor associated with it was inconclusive (see Section 3.3.2).

As discussed in Section 1.2.1, illusions of ungrammaticality are a critical test where representational and retrieval-based accounts diverge, and would indicate a contribution of feature distortion errors to attraction. Under representational approaches, feature distortion of the subject occurs before and regardless of the verb's features. Therefore, attraction should be observed on both ungrammatical and grammatical verbs, such that ungrammatical verbs tend to be erroneously perceived as grammatical and grammatical verbs as ungrammatical. Under pure retrieval approaches, attraction reflects conflicting retrieval cues of the verb. If all verbal cues correctly point to the subject, a distractor that mismatches the subject should not interfere. Our finding that attraction arises on grammatical verbs, as an illusion of ungrammaticality, therefore provides evidence for representational distortion accounts of attraction.

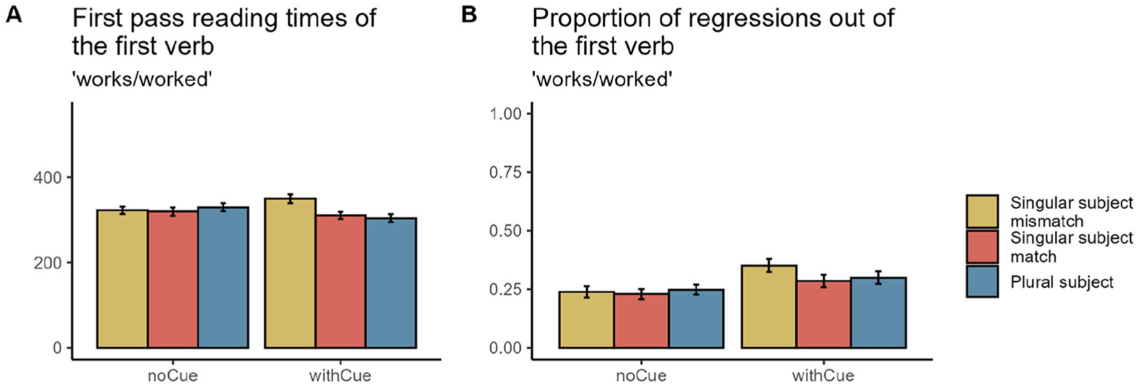

FIGURE 6
Mean first pass reading times **(A)** and mean proportion of regressions out **(B)** by condition, at the first verb. Error bars represent +/−SE.

We did not expect to observe an ungrammaticality illusion in our study, as this pattern has proven difficult to detect in previous research. However, the partial evidence for it could reflect a shift in reading strategies induced by our task. Participants were required to notice singular and plural features in every trial to complete the comprehension task. This could force deeper processing and reduce any ‘grammaticality’ biases, which can obscure ungrammaticality illusions (Hammerly et al., 2019; Laurinavichyute and von der Malsburg, 2024).

Agreement updating may partly rely on such response biases as well, and it could, therefore, trade off with the ungrammaticality illusion. This would explain why updating was not observed in this data set. Updating features of the subject based on the verb may only apply when comprehenders put higher faith in the form of a dependent (i.e., verbal agreement) than in their memory of the subject (see Keshev and Meltzer-Asscher, 2024, for a discussion of the rationality of an updating procedure). A bias for treating dependents as grammatical is also associated with the absence of ungrammaticality illusions, i.e., with the grammaticality asymmetry (Hammerly et al., 2019). It is possible that our secondary task made participants aware of agreement errors and undermined their trust in the grammaticality of the input. In this case, a verb-based update might be disfavored or undetectable in the task conditions where an illusion of ungrammaticality will arise.

4 Experiment 3a-b: SPR patterns partially compatible with task effects and feature updating

In Experiment 3, we implement two SPR sub-experiments: one (Experiment 3a) with 4-alternative forced-choice questions as we used in the eye tracking experiment, and the other (Experiment 3b) with yes/no questions. In Experiment 3a, we sought a conceptual replication of Experiment 2, using a similar task context but with a different reading measure. Experiment 3b is identical to Experiment 3a, except that the task context was changed to downplay the task relevance of singular/plural features. Broadly, we expected to see results that were more similar to the expected pattern when the secondary task involved answering yes/no questions, as these did not explicitly draw attention to the number features of the subject. By comparing the effect of the task on online reading measures, we aimed to examine agreement updating and its interaction with the task, which we speculated might have affected our eye tracking results.

4.1 Methods

4.1.1 Participants

We recruited participants until reaching a cap of 96 participants passing the exclusion criteria in each sub-experiment. We recruited a total of 292 self-reported native English-speaking participants through the Prolific Academic online platform. Of those, 166 were recruited for Experiment 3a and 126 for Experiment 3b, to accommodate different attrition rates across the sub-experiments. Participants gave informed consent and received monetary compensation of 6 USD for their participation (a rate of approximately 10 USD/h). This experiment was determined to be exempt research by the Institutional Review Board of the University of Massachusetts.

4.1.2 Materials

Items were the same as those of Experiment 2. In Experiment 3a, comprehension questions were identical to those used in Experiment 2. In Experiment 3b, we used yes/no comprehension questions as to the subject of the second verb (e.g., ‘Was the apprentice recruited by a top restaurant?’). Yes/no responses were balanced across items, with no questions targeting an interpretation where the distractor was the subject (e.g., Were the chefs recruited by a top restaurant?’). The experiments included the same fillers and catch trials as in Experiment 2.

4.1.3 Procedure

The experiment was a self-paced reading experiment. Each sentence was followed by one comprehension question. No feedback on response accuracy was provided. The experiment was

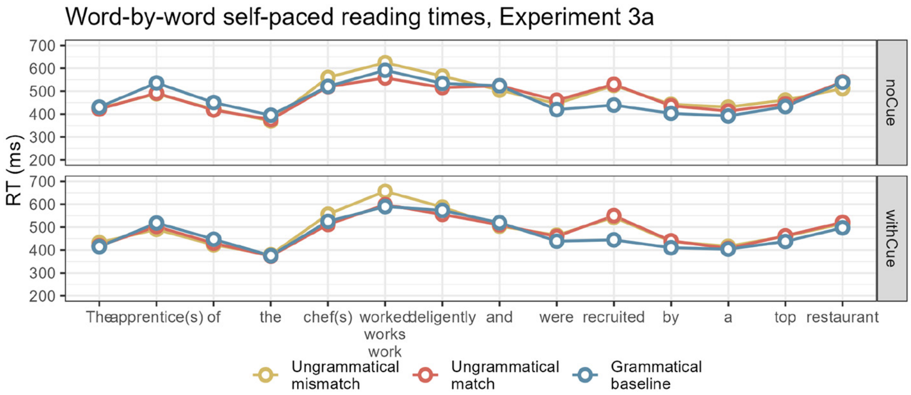

FIGURE 7
Word-by-word self-paced reading times by condition, Experiment 3a (4AFC questions).

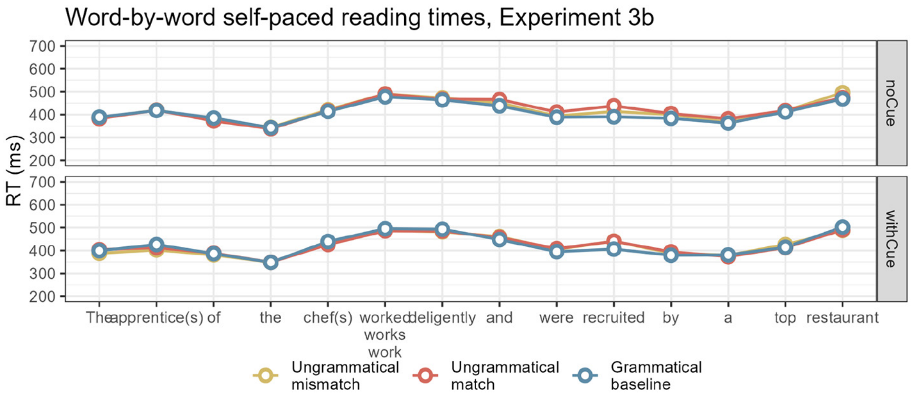

FIGURE 8
Word-by-word self-paced reading times by condition, Experiment 3b (yes/no questions).

implemented in PCIbex. Participants performed the experiment remotely on their own computer. Before starting the experiment, participants undertook a practice block of five sentences. The experiment took approximately 35 min.

4.1.4 Data analysis

Participants were excluded from the analysis if they failed more than one of the catch trials (69 participants in Experiment 3a, and 30 participants in Experiment 3b). For the remaining 193 participants (97 in Experiment 3a, and 96 participants in Experiment 3b), we excluded from analysis data points with response times below 100 ms and above 3,000 ms (affecting 2.09% of the data of Experiment 3a, and no data points in Experiment 3b). We analyze reading times on the second verb in separate models—one for the auxiliary carrying the number agreement, and one for the following past participle. Contrast coding and all modeling parameters are identical to those used for the reading time measures in Experiment 2. Results of the 4AFC task are analyzed separately in Section 5.

We also analyze RTs at the first verb and the subject, to probe for the ungrammaticality illusion effects observed in Experiment 2. However, it is important to consider that, in Experiment 2, the interaction effect revealing the ungrammaticality illusion (inhibitory attraction only in cases where the first verb carried overt agreement marking) only arose with measures that include re-reading (go-past reading times), which was not available in the self-paced reading paradigm.

4.2 Results

Reading times across regions are presented in Figures 7, 8. Model results for the critical region (the auxiliary verb) and the following word (the past participle) are summarized in Table 6.

In Experiment 3a, we observed only grammaticality effects, such that grammatical verbs were read faster than ungrammatical ones. These effects appeared at the spillover region, as a main effect and in the nested contrast. The models did not support a main effect of attraction, nested attraction contrasts, or an interaction between attraction and cue availability (Table 6).

In Experiment 3b, in addition to a grammaticality effect, the model also detected a main effect of attraction. Both attraction and grammaticality effects arose at the spillover and had the expected directionality: grammatical conditions were read faster than

TABLE 6 Results of Experiment 3.

Contrast label	Experiment *3a*		Experiment *3b*	
	Critical auxiliary	Past participle (spillover)	Critical auxiliary	Past participle (spillover)
M1: Cue availability	5 [−6, 17]	6 [−6, 19]	5 [−4, 13]	10 [1,19]
M1: Grammaticality	9 [−2, 21]	57 [44, 71]	6 [−2, 14]	24 [13, 34]
M1: Attraction	0 [−14, 15]	11 [−7, 28]	3 [−7, 13]	14 [3, 25]
M1 Interaction: Cue × Gram	−8 [−27, 11]	6 [−16, 28]	−7 [−25, 11]	−4 [−24, 15]
M1 Interaction: Cue × Attract	−18 [−42, 6]	−7 [−43, 31]	−7 [−28, 14]	−16 [−40, 8]
M2: Cue availability	5 [−6, 16]	6 [−6, 19]	4 [−4, 13]	10 [2, 18]
M2: No cue grammaticality	14 [−2, 28]	54 [38, 71]	10 [−1, 21]	26 [12, 40]
M2: With cue grammaticality	5 [−10, 20]	61 [42, 79]	2 [−10, 15]	22 [8, 35]
M2: No cue attraction	10 [−8, 28]	14 [−9, 37]	6 [−9, 22]	21 [5, 38]
M2: With cue attraction	−9 [−28, 11]	7 [−20, 35]	−1 [−15, 13]	6 [−10, 22]

Mean and 95% credible interval of the posterior distribution for fixed effects (on the ms scale). Credible intervals that do not cross zero are shaded gray. M1 is the main effects model. M2 is the nested contrasts model.

ungrammatical ones, and within the ungrammatical conditions, mismatch conditions were read faster than match conditions. An interaction between attraction and cue availability was not supported by the model. The credible interval crossed zero and there was only 91% chance of a directional effect [$\Pr(\beta < 0) = 0.91$].

However, the main effect of attraction seems to reflect a clear pairwise contrast only in no-cue conditions. The nested contrasts model detected attraction in the conditions where no agreement was available in the first conjunct (posterior mean [CrI]: 21 [5, 38]), but not in configurations with an intermediate agreement cue (posterior mean [CrI]: 6 [-10, 22]). It should also be noted that the model additionally detected a main effect of cue availability (posterior mean [CrI]: 10 [1, 19]), which could be partly driven by somewhat faster reading in the ungrammatical mismatch condition of with-cue sentences (see Figure 9).

Results at the first verb are analyzed in Appendix B. Neither sub-experiment replicated the ungrammaticality illusion observed in Experiment 2 in this region.

To further compare the effects between the two sub-experiments, we conducted a unified analysis where data from the spillover region of Experiment 3a and 3b were included. This analysis included the same predictors as the previous model in addition to a main effect of Experiment Format and its interaction with the other fixed effects. Model results are summarized in Table 7.

This analysis revealed a couple of contrasts between the two subexperiments. First, we observe a main effect of Experiment (posterior mean [CrI]: −65 [−102, −28]) such that RTs in sub-experiment 3a were longer than in sub-experiment 3b. Second, there was an interaction between Experiment and Grammaticality (posterior mean [CrI]: 34 [17, 50]) such that the grammaticality effect was reliably larger in Experiment 3a. The analysis additionally replicated the main effects of Grammaticality (observed in the separate analyses of each sub-experiment), Attraction, and Cue Availability (both observed in the separate analysis of Experiment 3b).

4.3 Discussion

The results of Experiment 3 provide only partial evidence in support of our original predictions. Specifically, the nested contrasts at the spillover region detected an attraction effect in no-cue conditions but not in with-cue conditions, in Experiment 3b. These findings are broadly consistent with the feature updating hypothesis. However, we failed to detect an interaction effect to corroborate that attraction was affected by our cue manipulation. Therefore, these results cannot be regarded as robust.

The results of Experiment 3 also confirm that the secondary task modulates the way in which agreement mismatches affect reading times. Participants seemed to be more sensitive to agreement in Experiment 3a than in Experiment 3b: Ungrammaticality incurred a large cost in Experiment 3a (Posterior mean [CrI]: 54 ms [38, 71]), twice the size and almost with no CrI overlap relative to this effect in Experiment 3b (Posterior mean [CrI]: 26 ms [12, 40]). The contrast was confirmed in a unified analysis that revealed a reliable interaction of Experiment and Grammaticality effect.

In Experiment 3a, where deeper processing and attention to number features were required for the comprehension task, we failed to observe agreement attraction and agreement updating. In Experiment 3b, on the other hand, we observed some evidence for both (a main effect of attraction and distinct patterns in pairwise contrasts). These contrasts were not supported by a compatible interaction (no evidence for interaction of Experiment with Attraction and/or with Cue availability). Therefore, we cannot derive strong conclusions from this pattern. However, it is worth mentioning that this numerical pattern is broadly similar to that observed by Parker (2019), who showed that attraction was more likely to arise in 'timed' as opposed to 'untimed' task contexts, and showed that this result naturally results from a sequential sampling of noisy memory contexts for more or less time. If one takes the suggestive by-subexperiment analysis to be informative (despite

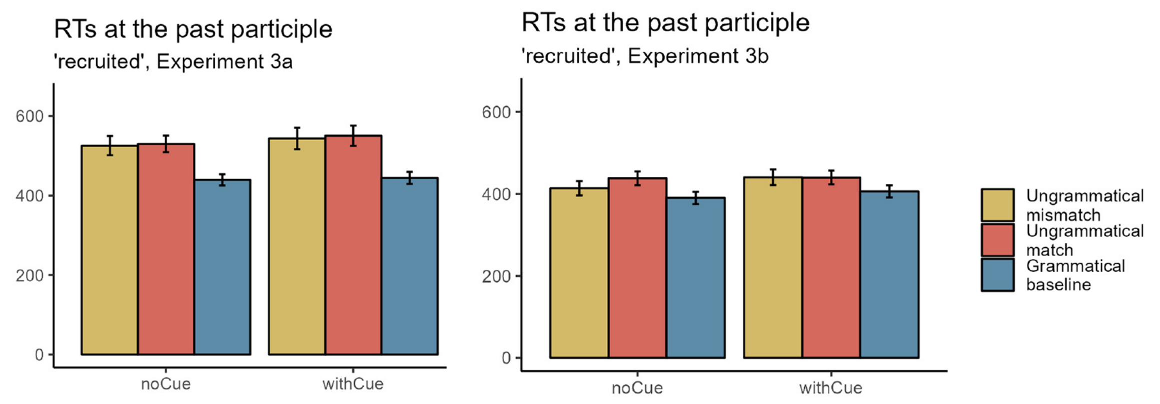

FIGURE 9
Mean self-paced reading times at the spillover region, across conditions and sub-experiments. Error bars represent +/−SE.

TABLE 7 Results of a unified analysis of Experiment 3.

Contrast label	Past participle (spillover)
Unified M1: Experiment	−65 [−102, −28]
Unified M1: Cue availability	8 [1, 15]
Unified M1: Grammaticality	41 [32, 49]
Unified M1: Attraction	12 [3, 22]
Unified M1 Interaction: Exp × Cue	−4 [−18, 11]
Unified M1 Interaction: Exp × Gram	34 [17, 50]
Unified M1 Interaction: Exp × Attract	−3 [−24, 18]
Unified M1 Interaction: Cue × Gram	1 [−13, 16]
Unified M1 Interaction: Cue × Attract	−11 [−32, 9]
Unified M1 Three-way Interaction: Experiment × Cue × Gram	10 [−18, 39]
Unified M1 Three-way Interaction: Experiment × Cue × Attract	9 [−39, 58]

Mean and 95% credible interval of the posterior distribution for fixed effects (on the ms scale). Credible intervals that do not cross zero are shaded gray.

the lack of interaction), our data reveals no attraction in the task context that generated longer reading times and hence more time for readers to gather memory samples in the service of processing the input (Parker, 2019).

The results of Experiment 3a also differ from those of Experiment 2, despite sharing the 4AFC task. First, the current experiment did not give rise to the same ungrammaticality illusion as in Experiment 2. This dovetails with the lack of a grammaticality illusion at the critical region in this experiment, and could be due to the same process. Therefore, we suggest that participants' awareness of agreement, coupled with the impossibility of rereading, might have shifted processing strategies such that attraction effects were small and undetectable. This aligns with a previous study that implemented a self-paced reading task with this kind of comprehension questions (Paape et al., 2021). It should also be mentioned that the ungrammaticality illusion in Experiment 2 was observed in go-past (regression path) times. It is possible that the unavailability of such rereading measures also made detection of this effect more difficult. However, it is also possible that the contrast between patterns in Experiment 3a and Experiment 2 indicates that the ungrammaticality illusion effect in Experiment 2 was spurious (i.e., reflects a Type I error).

5 Analysis of error patterns in Experiments 2–3: evidence for feature distortion and a pattern partially compatible with feature updating

Experiments 2 and 3a produced data on offline comprehension accuracy. Specifically, we obtained data about the representation that participants ended up with for the subject of the second conjunct. Participants had to choose the correct subject out of four alternatives—a plural and a singular version of the subject head and of the distractor noun. These data can provide insight into the way comprehenders interpret attraction configurations.[3]

Under Cue-Based Retrieval, comprehenders are expected to have veridical representations of the nouns in the sentence. Namely, they are not expected to select noun forms that did not appear in the sentence (e.g., 'apprentices' or 'chef' for 'the apprentice of the chefs'). Under this approach, errors in attraction configurations arise when comprehenders retrieve the wrong noun. Therefore, attraction configurations are expected to elicit more distractor responses ('chefs'). On the other hand, representational approaches suggest that access to the correct subject is retained (such that participants are unlikely to select 'chef' or 'chefs'). Attraction should

[3] We do not provide an analysis of the results of the yes/no comprehension questions in Experiment 3b. Such questions do not allow the full array of possible representations, and therefore, results might not reflect the underlying representations. In addition, accuracy in yes/no comprehension questions did not seem to differ between conditions, with a mean accuracy of 80%-84.5% in all conditions.

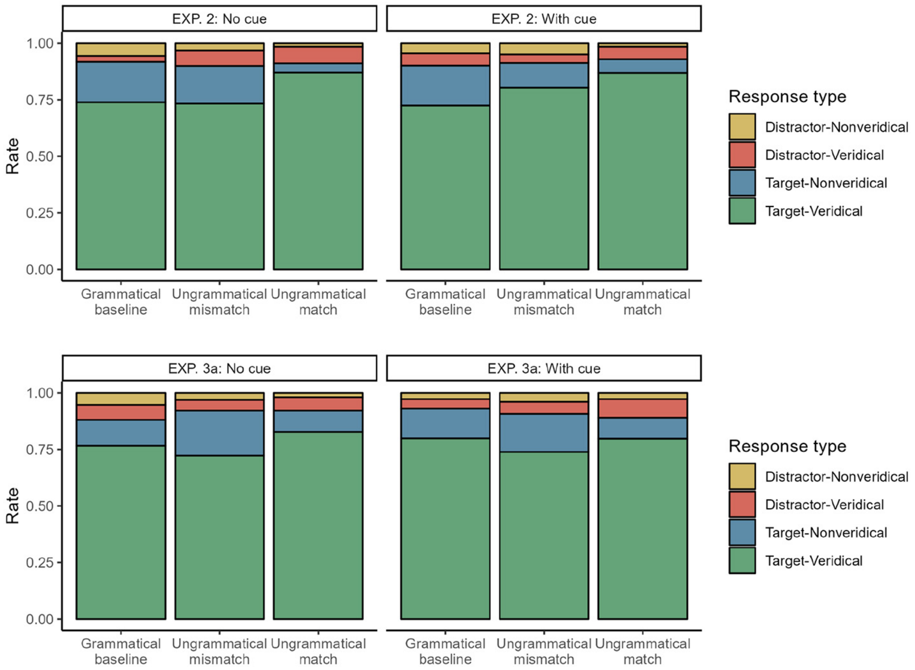

FIGURE 10
Mean response rates in a four-alternative forced-choice interpretation task. Break down across response types and conditions of Experiments 2 and 3a.

be reflected in increased rates of selecting a subject with a non-veridical number feature ('apprentices').

Here, we analyze the error patterns in the 4AFC task of Experiments 2 and 3. Figure 10 presents the distribution of responses across conditions. Table 8 provides the results of the statistical analyses. The dataset includes all participants who were not excluded for low accuracy in catch trials, and all data points from these participants. We analyze in separate models the effect of the experimental manipulation on the rate of selection of: the veridical subject (accuracy rate), the veridical distractor (distractor errors), and the non-veridical subject (number errors). We analyze these results with the same contrast coding as used for Experiments 2–3, in models with a binomial link function (with priors as in Experiment 1 and Experiment 2's regression rate analysis).

The results suggest that attraction is expressed in final interpretation as a lower rate of choosing the correct subject: accuracy rate in ungrammatical sentences was lower when the distractor mismatched the subject compared to when it matched it (attraction main effect on accuracy—Exp. 2: 0.84 [0.52, 1.16]; Exp. 3a: 0.61 [0.34, 0.89]). This reduction in accuracy is mirrored in an increase of number errors, as detected in a main effect of attraction on the rate of number errors (Exp. 2: −1.16 [−1.56, −0.75]; Exp. 3a: −94 [−1.25, −0.63]). The rate of distractor errors was not affected by the attraction manipulation; the posterior did not support an effect of attraction on the rate of distractor selection. If any trend is to be drawn from the data, it is in fact in the opposite direction, reflecting a numerically lower rate of distractor errors in attraction configurations. Thus, the results in both experiments suggest that attraction configurations encourage comprehenders to adopt a version of the subject noun that bears the wrong number marking, in line with representational approaches and in contrast with the predictions of Cue-Based Retrieval.

In Experiment 2, but not in Experiment 3a, there was also support for the feature updating hypothesis. The CrI for interaction between cue availability and attraction for rates of selecting the correct response was almost beyond zero (interaction effect on accuracy: −0.53 [−1.09, 0.03]), in a direction compatible with a larger effect of attraction when no agreement cue was available in the first conjunct. A more robust interaction was detected in the rates of non-veridical subject responses (cue-availability by attraction interaction on number errors: 0.98 [0.25, 1.74]). The increase in non-veridical subject responses, that is, number errors, in attraction configurations was more prominent when the first verb did not bear agreement marking (see Figure 10).

It should also be noted that another intriguing pattern emerged in the data of Experiment 2. We observe decreased accuracy in grammatical conditions compared to ungrammatical conditions (posterior mean [CrI]: 0.71 [0.27, 1.15]) and an increase in number errors for the same 'grammaticality' contrast (posterior mean [CrI]: −0.97 [−1.54, −0.41]). Since the grammatical conditions in our study involved plural subjects, this effect could reflect a bias for singular responses, which has also been observed in other studies (see Keshev et al., 2025; Keshev et al., under revision[4]).[5] The subject head was plural in grammatical conditions,

[4] Keshev, et al. (under revision). Feature distortion and memory updating: experimental and modeling evidence.

[5] Note that the direction of the grammaticality effect on error rate here contrasts with findings of Brehm et al. (2021) and Patson and Husband (2016) discussed in Section 1.2.2. In those studies ungrammaticality increased error

TABLE 8 Results of the final interpretation data from Experiments 2 and 3a.

Contrast label	Correct responses		Veridical distractor		Non-veridical target	
	Exp. 2	Exp. 3a	Exp. 2	Exp. 3a	Exp. 2	Exp. 3a
Cue availability	0.14 [−0.09, 0.36]	0.03 [−0.20, 0.24]	−0.14 [−0.58, 0.29]	−0.07 [−0.47, 0.31]	−0.06 [−0.33, 0.23]	−0.01 [−0.29, 0.27]
Grammaticality	0.71 [0.27, 1.15]	−0.17 [−0.66, 0.28]	0.48 [−0.08, 1.11]	0.22 [−0.25, 0.76]	−0.97 [−1.54, −0.41]	0.05 [−0.53, 0.64]
Attraction	0.84 [0.52, 1.16]	0.61 [0.34, 0.89]	0.34 [−0.24, 0.93]	0.28 [−0.28, 0.85]	−1.16 [−1.56, −0.75]	−0.94 [−1.25, −0.63]
Interaction: Cue × Gram	0.32 [−0.10, 0.72]	−0.22 [−0.69, 0.27]	−1.12 [−1.93, −0.27]	0.61 [−0.18, 1.38]	−0.06 [−0.58, 0.45]	−0.38 [−0.94, 0.16]
Interaction: Cue × Attract	−0.53 [−1.09, 0.03]	−0.23 [−0.75, 0.32]	0.34 [−0.51, 1.21]	0.23 [−0.56, 0.99]	0.98 [0.25, 1.74]	0.04 [−0.67, 0.70]

Mean and 95% credible interval (on the log-odd scale) of the posterior distribution for fixed effects. Credible intervals that do not cross zero are shaded gray.

but singular in ungrammatical conditions. It is possible that participants had a bias for the default, singular form, and that this caused them to misrepresent the subject's number features in the grammatical baseline.

6 General discussion

In this study, we set out to examine whether feature distortion contributes to interference or whether memory representations are fixed. We aimed to target this through the lens of feature updating. We hypothesized, following Keshev and Meltzer-Asscher (2024), that intermediate agreement marking (e.g., verbal agreement) would reduce vulnerability to attraction in later sites (e.g., a verb in a second conjunct). The resulting empirical pattern is complex, and not all results are in line with our predictions. Despite the substantial variability across experiments, we did nonetheless observe some evidence for updating effects. This effect was most evident in offline measures and secondary tasks: an interaction between attraction and the availability of an intermediate agreement cue in verb selection (Experiment 1), and in subject identification (Experiment 2, accuracy results). We did not see any clear evidence for this in online measures, perhaps because there was very little evidence for attraction to begin with in online measures (Experiment 2 and 3a). While the attraction patterns in reading times also seemed to diverge in cue and no-cue conditions in Experiment 3b, this pattern was not supported by a clear interaction effect (the 95% confidence interval included zero, but a 90% interval did not). It must be acknowledged that evidence from any single reading experiment is not very strong. Nonetheless, on balance, we find the footprint of feature updating across experiments—albeit with significant variability across measures which indicates that key features of this phenomenon, such as its interaction with the task, are not yet well understood.

We interpret these findings as broadly compatible with representational approaches to attraction with an additional rational updating function (Keshev and Meltzer-Asscher, 2024), and as evidence against the assumption that memory representations of the linguistic input are fixed. Modulation of attraction based on the availability of additional agreement marking is predicted if memory representations are associated with some degree of uncertainty, which can be influenced by other items in memory. Uncertainty in the memory encoding naturally permits updating as information becomes available (Xu and Futrell, 2025). Moreover, this updating should affect attraction rates only if attraction arises from uncertainty about the feature contents of representations. In retrieval-based models, attraction arises from activation of the wrong constituent. Intermediate (grammatical) agreement marking should not make the distractor a more or less prominent contestant for later retrieval. Therefore, evidence for updating supports representational approaches to attraction over retrieval-based models.

Is there a possible mechanism that could generate the effects of previous agreement cues in the cue-based retrieval framework, without adopting malleable or uncertain memory representations? It could be claimed that some activation level is allotted to each type of possible cue, even if it is not specified for the current retrieval trigger. In that case, activation would be a function of the absolute number of cue matches, rather than of the relative number of matches out of the relevant available cues. In the context of our study, such a system means that the lack of overt agreement marking on the verb deprives the subject of a potential activation boost, thus making its subsequent retrieval slower and more error-prone. However, this is not a common implementation of Cue-Based Retrieval. In most implementations, activation provided by each cue is scaled such that activation of only contextually relevant cues sums up to the maximal activation level. This current implementation is also more reasonable when considering the wide range of lexical-semantic cues that could in principle be used (Smith and Vasishth, 2020).

Our study also produced two other types of evidence for representational accounts of attraction. First, exploratory analyses in Experiment 2 revealed an illusion of ungrammaticality. This means that the verb in configurations like (9a) was read slowly relative to cases like (9b), as if it were ungrammatical. This effect

rates, which we interpreted as possible evidence for feature updating arising at the verb and distorting the number representations of the subject. In our design we can disentangle feature updating from grammaticality using the cue contrasts. Therefore the increased error rate in grammatical conditions does not bear on the updating question. The singular bias account fits with previous findings as in those studies the grammatical conditions included a singular subject with a singular verb rather than a baseline of grammatical plural agreement.

suggests attraction errors and sensitivity to mismatch with the distractor in grammatical verbs. The ungrammaticality illusion is not detected often, and it is task dependent (Hammerly et al., 2019; Laurinavichyute and von der Malsburg, 2024). However, its presence is a key prediction of representational approaches. On these approaches, a mismatching distractor distorts the number representation of the subject prior to the verb. Therefore, this distractor is expected to affect the processing of both grammatical and ungrammatical verbs (in opposite directions). Thus, the illusion of ungrammaticality contributes to our conclusion that memory representations are not fixed but vulnerable to distortions. However, we hasten to add that this effect was not replicated in Experiment 3a. Thus, although this effect is predicted by representational approaches, it did not appear consistently across experiments.

(9) a. The apprentice of the chefs works diligently.
b. The apprentice of the chef works diligently.

The finding that the verb in (9a) is more costly to process than in (9b) presents a challenge to pure retrieval-based accounts. The distractor in (9a) does not match the verb better than in (9b). In fact, the distractor matches more cues (i.e., the grammatical number cue) in (9b). Therefore, it should be more disruptive and hinder the retrieval of the subject in those match cases. Thus, Cue-Based Retrieval predicts a pattern opposite to the one we observed. Therefore, this finding presents a challenge to retrieval-based approaches.

Lastly, an analysis of error patterns also suggests that readers form non-veridical representations and points to a representational source of attraction. In Experiments 2 and 3a, a final interpretation probe required participants to recognize the correct subject. We found that attraction configurations (i.e., a mismatch between the subject and the distractor) were accompanied by increased rates of erroneously recalling the subject as plural. Rates of associating the distractor with the subject role were not modulated by verb-distractor match. This suggests that attraction involves forming an interpretation with a number error rather than one with the wrong subject noun. These results require further investigation due to the lack of online attraction effects and the possibility of late inference associated with ungrammaticality rather than memory distortion. Such additional examination, with rapid presentation and fully grammatical sentences, is available in Keshev et al. (see text footnote 4), who report similar patterns of feature distortion and updating.

Overall, across our experiments and dependent measures, we find several lines of evidence for representational models of interference. This points to a memory model which allows dynamic memory representations—representations that can be edited during maintenance in memory. This important consequence of our findings stands against assumptions of the prominent Cue-Based Retrieval model. Thus, we propose that a model of sentence processing has to incorporate representational shifts and interference to memory encodings. This could be implemented either in a model where interference arises only as representational distortion (Keshev et al., 2025) or a hybrid model where this arises in addition to retrieval errors (Yadav et al., 2023).

Still, each individual effect we report and interpret here was statistically weak (e.g., corroborated in a credible interval analysis but not in a Bayes Factor analysis or vice versa; or revealed only in nested comparisons but not in an interaction measure). Thus, if one does not take the sum of the detected patterns here to increase the reliability of each single effect, it is possible to interpret our data as consistent with no modulation of attraction and no ungrammaticality illusion. In that case, the only effect of interest that can be regarded as statistically robust is the occurrence of non-veridical responses to final comprehension questions for sentences with attraction. This mirrors previous findings, and poses a challenge to cue-based retrieval models (see Paape et al., 2021; Brehm et al., 2021; Keshev et al., see text footnote 4). While this pattern is reliable in our data, its interpretation is less clear: it has been debated whether final interpretation effects reflect the processing mechanisms responsible for dependency formation, or instead some post-interpretive misbinding (Dempsey et al., 2022). Thus, we do not regard the present evidence as decisively ruling out a pure cue-based retrieval account, since this can accurately predict the effects that were robust in multiple statistical tests.

6.1 Methodological implications

We interpret our results as broadly consistent with representational accounts of attraction that allow for uncertain associations between features and nominals in memory to be strengthened with additional evidence (e.g., Eberhard et al., 2005; Keshev et al., 2025). Still, we did not see our effects consistently across all experiments, suggesting that there is a complex interaction of secondary task and presentation modality (c.f. Hammerly et al., 2019; Laurinavichyute and von der Malsburg, 2024).

First, the ungrammaticality illusion was only observed in eyetracking while reading with a 4AFC comprehension task. In Experiment 2, using the eyetracking method, we observed an ungrammaticality illusion—a significant slowdown at the first verb site when the verb had agreement morphology, and when the target and distractor had mismatching number specifications. In Experiment 3, using self-paced reading, we failed to detect the ungrammaticality illusion on the first verb. The presence of an early RT effect of target-distractor mismatch is a key prediction of representational accounts, but it remains unclear why and how this effect interacts with the secondary task (Laurinavichyute and von der Malsburg, 2024). It is possible that the 4AFC task, which explicitly requires participants to distinguish singular and plural nouns, led comprehenders to prioritize resolving number information for noun phrases in the experiment. This would be analogous to how comprehension questions that target syntactic ambiguities modulate the processing of garden paths (Swets et al., 2008). It is also possible that the 4AFC task simply caused comprehenders to spend more time reading each word, which could modulate attraction rates if this means that comprehenders use the additional time this affords to gather more evidence from memory (Parker, 2019). In combination with the availability of rereading strategies, which eyetracking, but not self-paced reading, allows, this seems to produce the right conditions to observe the illusion of ungrammaticality.

A second methodological observation is that grammaticality effects and possibly attraction effects (illusion of grammaticality) as well are task-dependent. In Experiment 3a and Experiment 2, which included the 4AFC task, we saw a large effect of ungrammaticality (supported by an interaction of Experiment and Grammaticality) and little to no agreement attraction (not supported by the relevant interaction measure). It is possible that treating number features as a crucial comprehension task makes comprehenders more confident about the correct number of the subject and more surprised by any mismatch in the verb's number. This empirical pattern is broadly consistent with Paape et al. (2021), who used a 4AFC secondary task with a self-paced reading study on attraction inside RCs in Armenian. Paape et al. failed to see any agreement attraction in their reading time data, although their stimuli differ in a number of potentially important linguistic dimensions from our own, limiting how directly their results can be compared to our own.

The explanations, while plausible, are speculative in nature, since those task effects were not predicted. We believe that our results, in combination with Yadav et al. (2023) and Laurinavichyute and von der Malsburg (2024), support the broader conclusion that there is a complex interplay between the secondary task and the presence of classic 'agreement attraction' effects in reading time measures. However, to make sense of these patterns, explicit modeling is required: computationally implemented process models are required to explain why different tasks modulate the online pattern of attraction (as in Yadav et al., 2023).

6.2 Representational updating in other aspects of sentence processing

The current study was set to find evidence for updating of a word's feature representation in memory. We motivated this type of operation as a process which relies on grammatical knowledge and aims to minimize uncertainty about memory items. While results from this study are mixed, we would like to highlight that updating of individual memory items can be beneficial on other grounds as well. For example, during reanalysis of sentence structure (e.g., in Garden Path sentences or other ambiguities), position tagging and possibly other features (tense, part of speech, etc.) have to be modulated. Updating of transient feature values on previous memory items was also proposed in the past as a form of implementing relational cues in cue-based retrieval (Kush, 2013).

Similarly, discourse monitoring requires constant binding of new features to existing referents (Yu and Lau, 2023). While monitoring of discourse referents and events might use a slightly different working memory system than syntactic structure building, it is clear that the two interact. One example of this is the effect of discourse referents on agreement patterns. Morphologically singular nouns with notionally plural referents like collective nouns (e.g., *committee*) or phrases with a possible distributive reading (e.g., 'the label on the bottle') affect agreement patterns in a complex manner (Vigliocco et al., 1996; Humphreys and Bock, 2005; Smith et al., 2018; Sturt, 2022). Another example comes from the processing of temporarily ambiguous sentences. Updating discourse representations is often required when syntactic structure is updated, though lingering misinterpretations of Garden Path sentences reveal failures to this process (Christianson et al., 2001; Huang and Ferreira, 2021).

This is not to say that the fixed memory chunks model of cue-based retrieval is untenable, but it does suggest that any sentence processing model should incorporate the possibility of post-encoding editing of a memory items' features (see e.g., the ACT-R implementation of discourse representation theory in Brasoveanu and Dotlačil, 2020). This criticism also applies to representational sentence processing models (Keshev et al., 2025), which currently do not include controlled feature editing in line with updating needs but rather only accidental feature distortion.

More generally, the role of inferential processes and updating beliefs about sentence identity is central to the Noisy Channel model of sentence comprehension (e.g., Gibson et al., 2013). On this view of comprehension, language users draw inferences about likely meanings for the input using their knowledge of what is likely to have been said, and knowledge of what errors are likely to occur in transmission. This can lead language users to infer that the most likely meaning for the input is in fact a non-veridical interpretation of what was actually said, on the assumption that the intended meaning was corrupted by an error during transmission. This type of inference process can inform online sentence processing (Keshev and Meltzer-Asscher, 2021).

Here we have instead emphasized the role that an update function can play in minimizing uncertainty about the feature content of items in memory (see also Xu and Futrell, 2025), rather than noisy-channel inferences about the most likely meaning of the sentence. However, both approaches share the general perspective that comprehenders act to rationally offset the noise introduced in the comprehension process, either by external noise processes (as in the Noisy Channel model) or by internal error contributed by noisy memory systems (as in our feature updating approach and Xu and Futrell, 2025).

6.3 Conclusions

Sentence processing involves forming dependencies between incoming material and memory representations of past constituents. This crucial linguistic computation is vulnerable to memory interference, which can distort interpretation. We examined a prominent case study of interference—agreement attraction—in order to establish whether memory interference reflects only limits of the retrieval architecture or also distortion to memory representations themselves. We found different types of (weak) evidence for representational approaches to agreement attraction, including ungrammaticality illusion, interpretation errors, and feature updating at intermediate sites of the sentence (reducing vulnerability to attraction). Based on these findings, we propose that memory representations can be distorted and updated throughout sentence processing. Therefore, a dynamic model of feature editing and feature-based interference should be incorporated into models of memory processes in linguistic dependency formation.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/gdjne/.

Ethics statement

The studies involving humans were approved by the Institutional Review Board of the University of Massachusetts. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. MC: Conceptualization, Investigation, Methodology, Writing – review & editing. AM-A: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. BD: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was funded by the National Science Foundation (NSF-BSF #2146798 and NSF-BCS #1941485 to BD), the Binational Science Foundation (BSF #2021719 to AM-A), the Israeli Science Foundation (ISF #525/22 to AM-A), and the Azrieli Early Career Fellowship to MK.

Acknowledgments

We thank Kaiva Hinkle, Anissa Neal, Emi Aleksanyan, and David (Khiem) Le for their help in running Experiment 2.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2025.1708378/full#supplementary-material

References

Antón-Méndez, I., Nicol, J. L., and Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax* 5, 1–25. doi: 10.1111/1467-9612.00045

Arnett, N., and Wagers, M. (2017). Subject encodings and retrieval interference. *J. Mem. Lang.* 93, 22–54. doi: 10.1016/j.jml.2016.07.005

Avetisyan, S., Lago, S., and Vasishth, S. (2020). Does case marking affect agreement attraction in comprehension?. *J. Mem. Lang.* 112:104087. doi: 10.1016/j.jml.2020.104087

Badecker, W., and Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: evidence from gender and case in Slovak. *J. Mem. Lang.* 56, 65–85. doi: 10.1016/j.jml.2006.08.004

Bays, P., Schneegans, S., Ma, W. J., and Brady, T. (2022). Representation and computation in working memory. *PsyArXiv* [preprint]. doi: 10.31234/osf.io/kubr9

Bhatia, S., and Dillon, B. (2022). Processing agreement in Hindi: when agreement feeds attraction. *J. Mem. Lang.* 125:104322. doi: 10.1016/j.jml.2022.104322

Bleotu, A. C., and Dillon, B. (2024). Romanian (subject-like) DPs attract more than bare nouns: evidence from speeded continuations. *J. Mem. Lang.* 134:104445. doi: 10.1016/j.jml.2023.104445

Bock, K., and Cutting, J. C. (1992). Regulating mental energy: performance units in language production. *J. Mem. Lang.* 31, 99–127. doi: 10.1016/0749-596X(92)90007-K

Bock, K., and Eberhard, K. M. (1993). Meaning, sound and syntax in english number agreement. *Lang. Cogn. Process.* 8, 57–99. doi: 10.1080/01690969308406949

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., and Schriefers, H. (2001). Some attractions of verb agreement. *Cogn. Psychol.* 43, 83–128. doi: 10.1006/cogp.2001.0753

Bock, K., and Miller, C. A. (1991). Broken agreement. *Cogn. Psychol.* 23, 45–93. doi: 10.1016/0010-0285(91)90003-7

Brady, T. F., Konkle, T., and Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *J. Exp. Psychol. General* 138, 487–502. doi: 10.1037/a0016797

Brasoveanu, A., and Dotlačil, J. (2020). *Computational Cognitive Modeling and Linguistic Theory* (Singapore: Springer Nature), 294.

Brehm, L., Jackson, C. N., and Miller, K. L. (2019). Speaker-specific processing of anomalous utterances. *Q. J. Exp. Psychol.* 72, 764–778. doi: 10.1177/1747021818765547

Brehm, L., Jackson, C. N., and Miller, K. L. (2021). Probabilistic online processing of sentence anomalies. *Lang. Cogn. Neurosci.* 36, 959–983. doi: 10.1080/23273798.2021.1900579

Bürkner, P. C. (2017). brms: An R Package for Bayesian Multilevel Models using Stan. *J. Stat. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: a probabilistic programming language. *J. Stat. Softw.* 76, 1–32. doi: 10.18637/jss.v076.i01

Christianson, K., Hollingworth, A., Halliwell, J. F., and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cogn. Psychol.* 42, 368–407. doi: 10.1006/cogp.2001.0752

Cunnings, I., and Sturt, P. (2018). Retrieval interference and semantic interpretation. *J. Mem. Lang.* 102, 16–27. doi: 10.1016/j.jml.2018.05.001

Dempsey, J., Christianson, K., and Tanner, D. (2022). Misretrieval but not misrepresentation: a feature misbinding account of post-interpretive effects in number attraction. *Q. J. Exp. Psychol.* 75, 1727–1745. doi: 10.1177/17470218211061578

Deutsch, A., and Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: evidence from Hebrew. *J. Mem. Lang.* 60, 112–143. doi: 10.1016/j.jml.2008.07.001

Deutsch, A., and Dank, M. (2011). Symmetric and asymmetric patterns of attraction errors in producing subject–predicate agreement in Hebrew: an issue of morphological structure. *Lang. Cogn. Processes* 26, 24–46. doi: 10.1080/01690961003658420

Dillon, B. (2011). *Structured access in sentence comprehension* [Doctoral dissertation]. University of Maryland, College Park.

Dillon, B., Mishler, A., Sloggett, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *J. Mem. Lang.* 36, 147–164. doi: 10.1006/jmla.1996.2484

Eberhard, K. M., Cutting, J. C., and Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychol. Rev.* 112, 531–559. doi: 10.1037/0033-295X.112.3.531

Engelmann, F., Jäger, L. A., and Vasishth, S. (2019). The effect of prominence and cue association on retrieval processes: a computational account. *Cogn. Sci.* 43:12800. doi: 10.1111/cogs.12800

Ferreira, F., and Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Lang. Linguist. Compass* 1, 71–83. doi: 10.1111/j.1749-818X.2007.00007.x

Franck, J., Colonna, S., and Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Front. Psychol.* 6:807. doi: 10.3389/fpsyg.2015.00807

Franck, J., Vigliocco, G., and Nicol, J. (2002). Subject-verb agreement errors in French and English: the role of syntactic hierarchy. *Lang. Cogn. Process.* 17, 371–404. doi: 10.1080/01690960143000254

Futrell, R., Gibson, E., and Levy, R. P. (2020). Lossy-context surprisal: an information-theoretic model of memory effects in sentence processing. *Cogn. Sci.* 44:e12814. doi: 10.1111/cogs.12814

Gelman, A., Simpson, D., and Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy* 19:10. doi: 10.3390/e19100555

Gibson, E., Bergen, L., and Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proc. Natl. Acad. Sci.* 110, 8051–8056. doi: 10.1073/pnas.1216438110

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: an R package for estimating normalizing constants. *J. Stat. Softw.* 92, 1–29. doi: 10.18637/jss.v092.i10

Hammerly, C., Staub, A., and Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: experimental and modeling evidence. *Cogn. Psychol.* 110, 70–104. doi: 10.1016/j.cogpsych.2019.01.001

Hartsuiker, R. J., Schriefers, H. J., Bock, K., and Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Mem. Cogn.* 31, 1316–1326. doi: 10.3758/BF03195814

Haskell, T. R., Thornton, R., and MacDonald, M. C. (2010). Experience and grammatical agreement: statistical learning shapes number agreement production. *Cognition* 114, 151–164. doi: 10.1016/j.cognition.2009.08.017

Huang, Y., and Ferreira, F. (2021). What causes lingering misinterpretations of garden-path sentences: incorrect syntactic representations or fallible memory processes?. *J. Mem. Lang.* 121:104288. doi: 10.1016/j.jml.2021.104288

Humphreys, K. R., and Bock, K. (2005). Notional number agreement in English. *Psychon. Bull. Rev.* 12, 689–695. doi: 10.3758/BF03196759

Jäger, L. A., Engelmann, F., and Vasishth, S. (2017). Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis. *J. Mem. Lang.* 94, 316–339. doi: 10.1016/j.jml.2017.01.004

Jäger, L. A., Mertzen, D., Van Dyke, J. A., and Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: a large-sample study. *J. Mem. Lang.* 111:104063. doi: 10.1016/j.jml.2019.104063

Keshev, M., Cartner, M., Meltzer-Asscher, A., and Dillon, B. (2025). A working memory model of sentence processing as binding morphemes to syntactic positions. *Top. Cogn. Sci.* 17, 88–105. doi: 10.1111/tops.12780

Keshev, M., and Meltzer-Asscher, A. (2021). Noisy is better than rare: Comprehenders compromise subject-verb agreement to form more probable linguistic structures. *Cogn. Psychol.* 124:101359. doi: 10.1016/j.cogpsych.2020.101359

Keshev, M., and Meltzer-Asscher., A. (2024). The representation of agreement features in memory is updated during sentence processing: evidence from verb-reflexive interactions. *J. Mem. Lang.* 135:104495. doi: 10.1016/j.jml.2023.104495

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

Lago, S., Gračanin-Yuksek, M., Safak, D. F., Demir, O., Kirkici, B., and Felser, C. (2019). Straight from the horse's mouth: agreement attraction effects with Turkish possessors. *Linguist. Approach. Bilingualism* 9, 398–426. doi: 10.1075/lab.17019.lag

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., and Phillips, C. (2015). Agreement attraction in Spanish comprehension. *J. Mem. Lang.* 82, 133–149. doi: 10.1016/j.jml.2015.02.002

Laurinavichyute, A., and von der Malsburg, T. (2024). Agreement attraction in grammatical sentences and the role of the task. *J. Mem. Lang.* 137:104525. doi: 10.1016/j.jml.2024.104525

Lee, M. D., and Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *J. Multivariate Anal.* 100, 1989–2001. doi: 10.1016/j.jmva.2009.04.008

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

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

Lewis, S., and Phillips, C. (2015). Aligning grammatical theories and language processing models. *J. Psycholinguist. Res.* 44, 27–46. doi: 10.1007/s10936-014-9329-z

Molinaro, N., Kim, A., Vespignani, F., and Job, R. (2008). Anaphoric agreement violation: an ERP analysis of its interpretation. *Cognition* 106, 963–974. doi: 10.1016/j.cognition.2007.03.006

Nicenboim, B., and Vasishth, S. (2018). Models of retrieval in sentence comprehension: a computational evaluation using Bayesian hierarchical modeling. *J. Mem. Lang.* 99, 1–34. doi: 10.1016/j.jml.2017.08.004

Norris, D., and Kalm, K. (2021). Chunking and data compression in verbal short-term memory. *Cognition* 208:104534. doi: 10.1016/j.cognition.2020.104534

Paape, D., Avetisyan, S., Lago, S., and Vasishth, S. (2021). Modeling misretrieval and feature substitution in agreement attraction: a computational evaluation. *Cogn. Sci.* 45:e13019. doi: 10.1111/cogs.13019

Parker, D. (2019). Two minds are not always better than one: modeling evidence for a single sentence analyzer. *Glossa J. General Linguist.* 4:64. doi: 10.5334/gjgl.766

Patson, N. D., and Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Q. J. Exp. Psychol.* 69, 950–971. doi: 10.1080/17470218.2014.992445

Pearlmutter, N. J., Garnsey, S. M., and Bock, K. (1999). Agreement processes in sentence comprehension. *J. Mem. Lang.* 41, 427–456. doi: 10.1006/jmla.1999.2653

Phillips, C., Wagers, M. W., and Lau, E. F. (2011). *5: Grammatical Illusions and Selective Fallibility in Real-Time Language Comprehension*. Leiden: Brill.

R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (Accessed June 26, 2024).

Raab, D. (1962). Statistical facilitation of simple reaction times. *Transac. N. Y. Acad. Sci.* 24, 574–590. doi: 10.1111/j.2164-0947.1962.tb01433.x

Schad, D. J., Vasishth, S., Hohenstein, S., and Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: a tutorial. *J. Mem. Lang.* 110:104038. doi: 10.1016/j.jml.2019.104038

Slioussar, N. (2018). Forms and features: the role of syncretism in number agreement attraction. *J. Mem. Lang.* 101, 51–63. doi: 10.1016/j.jml.2018.03.006

Slioussar, N., and Malko, A. (2016). Gender agreement attraction in Russian: production and comprehension evidence. *Front. Psychol.* 7:1651. doi: 10.3389/fpsyg.2016.01651

Smith, G., Franck, J., and Tabor, W. (2018). A self-organizing approach to subject–verb number agreement. *Cogn. Sci.* 42, 1043–1074. doi: 10.1111/cogs.12591

Smith, G., and Vasishth, S. (2020). A principled approach to feature selection in models of sentence processing. *Cogn. Sci.* 44:12918. doi: 10.1111/cogs.12918

Staub, A. (2009). On the interpretation of the number attraction effect: response time evidence. *J. Mem. Lang.* 60, 308–327. doi: 10.1016/j.jml.2008.11.002

Sturt, P. (2022). Syntactic and semantic mismatches in English number agreement. *Glossa Psycholinguist.* 1, 1–33. doi: 10.5070/G6011159

Swets, B., Desmet, T., Clifton, C., and Ferreira, F. (2008). Underspecification of syntactic ambiguities: evidence from self-paced reading. *Mem. Cogn.* 36, 201–216. doi: 10.3758/MC.36.1.201

Tanner, D., Nicol, J., and Brehm, L. (2014). The time-course of feature interference in agreement comprehension: multiple mechanisms and asymmetrical attraction. *J. Mem. Lang.* 76, 195–215. doi: 10.1016/j.jml.2014.07.003

Tucker, M. A., Idrissi, A., and Almeida, D. (2015). Representing number in the real-time processing of agreement: self-paced reading evidence from Arabic. *Front. Psychol.* 6:347. doi: 10.3389/fpsyg.2015.00347

Tucker, M. A., Idrissi, A., and Almeida, D. (2021). Attraction effects for verbal gender and number are similar but not identical: self-paced reading evidence from Modern Standard Arabic. *Front. Psychol.* 11:586464. doi: 10.3389/fpsyg.2020.586464

Türk, U., and Logačev, P. (2024). Agreement attraction in Turkish: the case of genitive attractors. *Lang. Cogn. Neurosci.* 39, 448–454. doi: 10.1080/23273798.2024.2324766

Vasishth, S., Brüssow, S., Lewis, R. L., and Drenhaus, H. (2008). Processing polarity: how the ungrammatical intrudes on the grammatical. *Cogn. Sci.* 32, 685–712. doi: 10.1080/03640210802066865

Vasishth, S., Nicenboim, B., Engelmann, F., and Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends Cogn. Sci.* 23, 968–982. doi: 10.1016/j.tics.2019.09.003

Vigliocco, G., Butterworth, B., and Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: differences in the role of conceptual constraints. *Cognition* 61, 261–298. doi: 10.1016/S0010-0277(96)00713-5

Vigliocco, G., Butterworth, B., and Semenza, C. (1995). Constructing subject-verb agreement in speech: the role of semantic and morphological factors. *J. Mem. Lang.* 34, 186–215. doi: 10.1006/jmla.1995.1009

Vigliocco, G., and Franck, J. (1999). When sex and syntax go hand in hand: gender agreement in language production. *J. Mem. Lang.* 40, 455–478. doi: 10.1006/jmla.1998.2624

Vigliocco, G., and Franck, J. (2001). When sex affects syntax: contextual influences in sentence production. *J. Mem. Lang.* 45, 368–390. doi: 10.1006/jmla.2000.2774

Vigliocco, G., and Nicol, J. (1998). Separating hierarchical relations and word order in language production: is proximity concord syntactic or linear?. *Cognition* 68, B13–B29. doi: 10.1016/S0010-0277(98)00041-9

Wagers, M., Lau, E. F., and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *J. Mem. Lang.* 61, 206–237. doi: 10.1016/j.jml.2009.04.002

Xu, W., and Futrell, R. (2025). Informativity enhances memory robustness against interference in sentence comprehension. *J. Mem. Lang.* 142:104603. doi: 10.1016/j.jml.2024.104603

Yadav, H., Paape, D., Smith, G., Dillon, B. W., and Vasishth, S. (2022). Individual differences in cue weighting in sentence comprehension: an evaluation using approximate Bayesian computation. *Open Mind* 6, 1–24. doi: 10.1162/opmi_a_00052

Yadav, H., Smith, G., Reich, S., and Vasishth, S. (2023). Number feature distortion modulates cue-based retrieval in reading. *J. Mem. Lang.* 129:104400. doi: 10.1016/j.jml.2022.104400

Yu, X., and Lau, E. (2023). The binding problem 2.0: beyond perceptual features. *Cogn. Sci.* 47:e13244. doi: 10.1111/cogs.13244

Zehr, J., and Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. doi: 10.17605/OSF.IO/MD832