COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 48 (2024) e13497

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13497

Incremental Discourse-Update Constrains Number Agreement Attraction Effect

Sanghee J. Kim, Ming Xiang

Department of Linguistics, University of Chicago

Received 8 September 2023; received in revised form 13 June 2024; accepted 22 August 2024

Abstract

While a large body of work in sentence comprehension has explored how different types of linguistic information are used to guide syntactic parsing, less is known about the effect of discourse structure. This study investigates this question, focusing on the main and subordinate discourse contrast manifested in the distinction between restrictive relative clauses (RRCs) and appositive relative clauses (ARCs) in American English. In three self-paced reading experiments, we examined whether both RRCs and ARCs interfere with the matrix clause content and give rise to the agreement attraction effect. While the standard attraction effect was consistently observed in the baseline RRC structures, the effect varied in the ARC structures. These results collectively suggest that discourse structure indeed constrains syntactic dependency resolution. Most importantly, we argue that what is at stake is not the static discourse structure properties at the global sentence level. Instead, attention should be given to the incremental update of the discourse structure in terms of which *discourse questions* are active at any given moment of a discourse. The current findings have implications for understanding the way discourse structure, specifically the active state of discourse questions, constrains memory retrieval.

Keywords: Sentence comprehension; Discourse structure; Appositive relative clause; Restrictive relative clause; Number agreement attraction; Memory retrieval

Correspondence should be sent to Sanghee J. Kim, Department of Linguistics, University of Chicago, 1115 E 58th St, Room 203, Chicago, IL 60637, USA. E-mail: sangheekim@uchicago.edu

1. Introduction

A large body of work in sentence comprehension has centered on questions about parsing, that is, how the syntactic structure of a sentence is established in real time during language comprehension. A particularly fruitful empirical domain for this investigation involves investigating how nonlocal linguistic dependencies, such as subject-verb agreement, anaphoric, or *wh*-filler-gap dependences, are resolved. Establishing linguistic dependencies relies on the successful identification and retrieval of the elements in the dependency chain. Various types of information have been discussed as useful cues that could facilitate the process of dependency building, including formal morpho-syntactic features of the dependency elements, such as number and gender features (e.g., Slioussar & Malko, 2016), lexical semantic information (e.g., Kwon, Ong, Chen, & Zhang, 2019), or world knowledge regarding the plausibility of a given event (e.g., Cunnings & Sturt, 2018).

Despite the extensive amount of work on different types of information employed for dependency resolution, it is less clear how information related to *discourse* affects the syntactic process of dependency building. Existing work concerning the effect of discourse on language processing has mostly examined how discourse salience and focus in information structure affects language comprehension (Cutler & Fodor, 1979; Clifton & Frazier, 2018; Foraker & McElree, 2007; Sanford & Sturt, 2002; Sturt, Sanford, Stewart, & Dawydiak, 2004). These studies have focused on examining whether linguistic elements that receive greater attention and salience have a more robust representation in memory and whether they are retrieved and recognized faster. However, what is less explored is whether discourse information exerts a direct influence on syntactic parsing (e.g., Frazier & Clifton, 2005). The current study examines this question.

The theoretical notion of discourse can be used to refer to a variety of different aspects of discourse context, including information packaging (e.g., given vs. new information) (Birner & Ward, 1998; Kaiser & Trueswell, 2004), strategies that interlocutors engage in to keep track of conversation goals and their relevant conversational moves to achieve the goal (Bruce & Farkas, 2007; Farkas & Bruce, 2010; Lewis, 1979; Roberts, 2004), or coherence relationship between discourse units (Hobbs, 1985; Kehler, 2002). In the current work, we consider discourse to be a structured representation composed of a sequence of discourse units (words and clauses) and their relations to one another (as in Asher & Lascarides, 2003; Polanyi, 1988). We specifically focus on the contrast between *main discourse* versus *subordinate discourse* information. The particular linguistic constructions we examine involve appositive relative clauses (ARCs) and restrictive relative clauses (RRCs), see an example in (1). As we will introduce in detail in the next section, information hosted in an ARC is usually assumed to belong to the subordinate discourse, whereas information hosted by RRCs is part of the main discourse.[1]

(1) Appositive relative clause (ARC) and restrictive relative clause (RRC)
   a. The waitress, **who sat near the girl**, was unhappy. [ARC]
   b. The waitress **who sat near the girl** was unhappy. [RRC]

Our primary interest is whether and how the different discourse status of linguistic information—either being part of the main or subordinate discourse structure—is used to constrain the dependency-building process. To this end, we investigate the subject-verb agreement dependency resolution, specifically the number agreement attraction effect (e.g., Wagers, Lau, & Phillips, 2009). Agreement attraction has become a hallmark case that demonstrates the effect of a cue-based memory retrieval mechanism. To the extent that our findings reveal the effects of discourse structure on the agreement attraction effect, this has implications about how discourse information constrains memory retrieval.

A total of three self-paced reading experiments will be reported. Our results suggest that discourse information, specifically the distinction between main and subordinate discourse status, plays an important role in constraining the agreement attraction effect. Interestingly, however, the static division between main versus subordinate structures alone is not sufficient to account for all the results. Instead, as we will argue, what is at stake is the incremental update of the discourse structure, guided by what we term the *active state of discourse questions*.

2. Structuring discourse

2.1. A static division at the global discourse level

Discourse information is hierarchically structured. One of the key structural distinctions in discourse is between discourse units that are part of the *main discourse* and discourse units that form the *subordinate discourse* (subordinated to the main units) (Asher & Lascarides, 2003; Hunter & Asher, 2016; Jasinskaja, 2016). Sentences containing ARCs and RRCs exhibit this contrast. For instance, in (2a), two pieces of information are expressed in this example: *the waitress was unhappy* and *the waitress sat near the girl*. The first piece of information contributes to the main point of the utterance (also known to be *at-issue*), whereas the latter piece of information describes additional, supplementary information about the waitress (AnderBois, Brasoveanu, & Henderson, 2015; Murray, 2014; Potts, 2005). However, the same information that "the waitress sat near the girl" becomes part of the main point when it is situated inside an RRC (2b), by virtue of the fact that the RRC serves the function of restricting the specific referent.

(2) Two types of RCs and their discourse structure
  a. ARC is part of subordinate discourse
     The waitress, **who sat near the girl**, was unhappy.
  b. RRC is part of main discourse
     The waitress **who sat near the girl** was unhappy.

There is a large body of work discussing the semantic distinction between an ARC and the main assertion of an utterance (Bach, 1999; Dever, 2001; Potts, 2005). Additionally, appositives have also been argued to perform speech acts independently from the hosting matrix clause (Frazier, Dillon, & Clifton, 2018; Koev, 2022), or demonstrate distinct prosodic con-

tours (Dehé & Kavalova, 2007; Truckenbrodt, 2015). Some researchers have characterized ARCs as syntactic "orphans" (Haegeman, 2008), treating them to be syntactically independent from the matrix clause (Canac-Marquis & Tremblay, 1998; Emonds, 1979; Fabb, 1990; McCawley, 1982; Ott, 2016; Onea & Ott, 2022; Ross, 1967; Safir, 1986).[2] In the current study, drawing from discourse theories such as Segmented Discourse Represented Theory (Asher, 1993; Asher & Lascarides, 2003), we adopt the distinction between main and subordinate discourse units to understand the discourse status of ARC and RRC content. From this view, ARCs contribute to subordinate discourse information (+SUBORDINATE), whereas RRCs are part of main discourse information (+MAIN).

Some previous experimental work has found evidence that main and subordinate discourse information appears not to interact with each other during processing (Dillon, Clifton, & Frazier, 2014; Kroll & Wagers, 2019; McInnerney & Atkinson, 2020). Dillon et al. (2014), for example, looked at the differences between the two conditions in (3), comparing appositives with RRC structures.

(3) Experimental material in Dillon et al. (2014)
   a. That butcher **who was in the busy shop (Amy visited on Third Avenue)** bought his meat from local farmers. [RRC] (Longer clause in parenthesis)
   b. That butcher, **the one in the busy shop (Amy visited on Third Avenue)**, bought his meat from local farmers. [Appositive] (Longer clause in parenthesis)

In their acceptability rating task, the authors identified a length penalty effect in the baseline RRC condition (3a), where longer intervening RRC structures incurred more processing burden compared to shorter ones. This aligns with previous research demonstrating processing costs associated with longer embedded clauses (e.g., Gibson, 1998, 2000; Grodner & Gibson, 2005; Hale, 2001; Lewis & Vasishth, 2005; Levy, 2008; McElree, Foraker, & Dyer, 2003; Van Dyke & Lewis, 2003; Warren & Gibson, 2002). However, the length penalty effect in the RRC condition was reduced in the appositive condition (3b). Similar findings regarding the reduced length penalty with appositives were also reported in Kroll and Wagers (2019) and Duff, Anand, Brasoveanu, and Rysling (2023).

These earlier findings showed that language comprehension is sensitive to the distinction of main versus subordinate discourse status, raising the possibility that main and subordinate discourse information might be managed separately, potentially deploying different working memory spaces (Dillon et al., 2014). This hypothesis, however, turns out to be too strong in light of some later findings, which we turn to below.

2.2. Challenges to the static division view

Some later studies, while replicating the finding that linguistic information in main and subordinate discourse do not interact, also discovered evidence suggesting that the two types of information are not entirely separated. For example, Dillon, Clifton, Sloggett, and Frazier (2017) examined ARC and RRC processing using a *wh*-filler-gap dependency (4). The experimental items involved a *wh*-filler (who_1) connected to a sentence-final gap (*dinner for* ___).

Additionally, there was an intervening *wh*-dependency (*who*$_2$ *bought Italian ham*), varied by the type of RC structure, either an RRC or an ARC.

(4) Experimental material in Dillon et al. (2017)
   a. The butcher asked *who*$_1$ the lady ***who***$_2$ bought Italian ham was cooking *dinner for* ___. [RRC]
   b. The butcher asked *who*$_1$ the lady, ***who***$_2$ bought Italian ham, was cooking *dinner for* ___. [ARC]

Two notable findings were reported in their eye-tracking experiments. First, a processing cost was observed at the gap position (*dinner for*) in the RRC condition compared to its less complex baseline control without a *wh*-filler-gap dependency (e.g., "The butcher asked *if* the lady(,) who…"), an expected complexity effect (Frazier & Clifton, 1989; Gordon, Hendrick, & Johnson, 2001; Gibson & Warren, 2004; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005; Staub, 2010; Wagers & Phillips, 2014). The corresponding complexity effect, however, did not arise in the ARC condition, replicating general findings from Dillon et al. (2014). Second, a crucial finding was that no differences were found between ARC and RRC conditions in the intervening RC regions (*who*$_2$…), with both conditions showing a reading time slowdown, likely due to interference from an open outer *wh*-dependency (*who*$_1$…).

The empirical generalization so far appears to be the following: the subordinate discourse information encoded by an ARC can interact with the main discourse information encoded in the main clause. However, the interaction is only present when the subordinate discourse unit, that is, ARC, has not been closed off. Once the parser progresses beyond the right boundary of the subordinate discourse unit (e.g., beyond the RC), the information within the subordinate discourse unit becomes more separated from the main discourse unit. Dillon et al. (2017) propose that the rapid structural decay of ARC content after processing it could be a contributing factor, an idea also argued by Duff et al. (2023). They posit that content within the subordinate discourse structure after the RC boundary becomes "discounted," exerting "less influence on downstream parsing […] than other material." We will revisit this proposal about structural decay along with a comparison with a few alternative approaches in the General Discussion section.

3. The current study

3.1. Incremental construction of discourse questions

To capture both the separation and interaction between main and subordinate discourse units, we propose a **question-based approach**, inspired by the view that discourse can be represented as a hierarchical set of questions and corresponding answers (Büring, 2003; Roberts, 2012). In line with the incremental discourse construction framework (Asher & Lascarides, 2003; Jasinskaja, 2016; Lascarides & Asher, 2008; Riester, 2019), our proposals adhere to two fundamental principles. First, each discourse unit, whether main or subordinate, can intro-

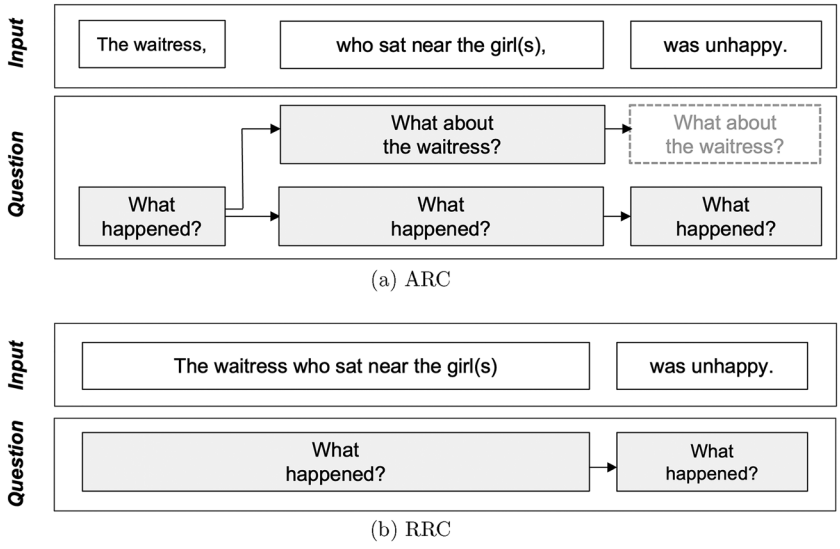


Fig. 1. Incremental construction (from left to right) of discourse questions. The box with dotted lines indicates the discourse question is no longer active.

duce a new *discourse question*.[3] Second, a discourse question remains in an *active state* until resolved; and after the resolution of a discourse question, it will be removed (or *popped off*) from the question stack. The extent to which the main and subordinate discourse units can interact with each other is determined by whether the discourse questions hosted by them are still in an active state.

Fig. 1 outlines the incremental generation and removal of discourse questions while parsing sentences with an ARC and an RRC. The top *input* panel presents incoming information, while the bottom *question* panel depicts the relevant discourse questions. In the case of an ARC (Fig. 1a), when the subject noun phrase (*The waitress*) is introduced, a discourse question like "What happened to the waitress?" could be raised, remaining active until resolved by the matrix clause (*the waitress was unhappy*). Upon encountering the ARC (*who sat near the girl*), a new discourse question ("What about the waitress?") is added, persisting until the ARC concludes, and was subsequently resolved. Once this question is resolved, it is popped off (i.e., removed) from the question stack and is deactivated, leaving only the main question ("What happened?") active in the question panel.

For RRCs, we consider them as intersective modifiers (Kratzer & Heim, 1998; Partee, 1975), being integrated into the larger discourse unit to which the RRC is attached. As shown in Fig. 1b, the RRC contributes to the main discourse question ("What happened?") without initiating an independent discourse question. While RRCs might prompt subquestions (e.g., "Which waitress?"), these are encompassed within the overarching super-question ("What happened?") due to their restrictive function, aligning with observations in Göbel (2019) that RRCs may function less as a separate discourse unit than ARCs. The super-question associated with the main discourse unit persists until it is resolved at the end of the sentence.

We further assume that the linguistic content that is hosted within the active questions is the most salient and attended component in the ongoing discourse, and that information hosted under resolved questions is less salient (similar to Grosz & Sidner, 1986). Crucially, the incremental update of which discourse questions are active or inactive creates different kinds of *salience domains* that could have consequences for information access and retrieval. For instance, in Fig. 1a, when both discourse questions ("What about the waitress?" and "What happened?") are active, that is, while the parser is still processing the ARC content, the linguistic information associated with those questions is highly salient in the discourse. That is to say, both the main clause content and the ARC clause content would be subject to memory access and retrieval. It is, therefore, possible that information from these two types of clauses could interfere with each other. However, once the parser moves beyond the right boundary of the ARC, resolving and removing the ARC-associated question from the stack, only the question associated with the main clause ("What happened?") remains active. At this point, the main clause content and the ARC clause content belong to different salience domains, with the information in the ARC clause less accessible and less likely to interfere with the main clause content.

3.2. Number agreement attraction effect with ARCs and RRCs

The current study uses the well-studied agreement attraction effect as a testing case to examine whether the incremental update of discourse questions has consequences on information access and retrieval. Agreement attraction is a hallmark phenomenon that demonstrates the cue-based memory retrieval mechanisms in sentence processing. In (5), although both (5a) and (5b) exhibit ungrammatical number agreement on the verb (*were*), (5b) generally displays higher acceptability ratings and shorter reading times on the verb compared to (5a), showing a standard attraction effect. This effect is typically explained as a misretrieval of the intervening plural distractor noun phrase (*the girls*) in (5b), due to the partial feature match ([+PLURAL]) between the verb (*were*) and the intervening distractor (*the girls*) (e.g., Parker & An, 2018; Wagers et al., 2009).[4]

(5) Experimental material in Parker and An (2018)
   a. *The waitress **who sat near** *the girl* surprisingly were unhappy about the noise. [Singular distractor]
   b. *The waitress **who sat near** *the girls* surprisingly were unhappy about the noise. [Plural distractor]

We use the number agreement attraction effect as a probe to examine the effect of incremental updates of discourse questions. The key comparison we focus on is as follows:

(6) a. *The waitress **who sat near** *the girls* surprisingly were unhappy about the noise. [RRC]
   b. *The waitress, **who sat near** *the girls*, surprisingly were unhappy about the noise. [ARC]

Both sentences in (6) are ungrammatical. (6a) represents the baseline condition where the number agreement attraction effect has been observed in prior studies (e.g., Parker and An, 2018). In contrast, (6b) is almost identical to the baseline, except that the distractor noun (*the girls*) is part of a subordinate ARC structure. Additionally, the retrieval site (*were*) appears after the ARC boundary is closed off. Based on the discourse-question approach outlined in the previous section, the ARC-associated content becomes less accessible at the retrieval site, due to the fact that the discourse question associated with the ARC has been resolved and deactivated. This predicts that information hosted under the ARC, for example, the distractor noun phrase *the girls*, is less likely to interfere with information hosted in the main clause, leading to an absence of (or at least reduced) agreement attraction effect in (6b), unlike its RRC counterpart (6a). Previous studies on similar constructions found some evidence in support of this prediction, but the results were also mixed (McInnerney and Atkinson, 2020; Ng & Husband, 2017).

In Experiment 1, we establish the core contrast between ARC and RRC. To preview the findings, we observe a standard agreement attraction effect in the RRC conditions but not in the ARCs, consistent with the discourse-question-based proposal. To further examine this proposal, in Experiments 2 and 3, we test situations in which the retrieval of an agreement controller is initiated while the discourse question associated with the ARC content is still active. In these configurations, we find a standard number agreement attraction effect in the ARC conditions, just like in the RRC conditions.

4. Experiment 1

4.1. Methods

4.1.1. Subjects

We recruited 120 native speakers of American English residing in the United States via the Prolific platform. Participation recruitment was conducted with IRB approval from the local institution. Two participants were excluded from the main analysis as they self-reported that their first language was not English, leaving us 118 participants for the analysis (mean age = 30.94; range: 18–50). The duration of the experiment was approximately 20 min, and participants were paid 3.50 USD in compensation.

4.1.2. Materials

Material for Experiment 1 is presented in Table 1. The material consisted of 48 items in a fully crossed 2 x 2 x 2 design with grammaticality (grammatical vs. ungrammatical), distractor number (singular vs. plural), and clause type (ARC vs. RRC) as factors. The material with RRCs was adapted from Experiment 1 in Parker and An (2018). All the target sentences had the structure of *NP1(,) who VERB Preposition NP2(,) ADVERB {was/were}…* (e.g., "The waitress(,) who sat near the girl(s)(,) unsurprisingly was/were unhappy about all the noise."). NP1, the target subject, was always a singular noun (e.g., *The waitress*). The main auxiliary verb was varied by grammaticality: *was* (grammatical) or *were* (ungrammatical). The

Table 1
Sample set of experimental items for Experiment 1

Condition	Sentence
ARC-Sg-Gr	The waitress, / who / sat / near / *the girl*, / unsurprisingly / **was** / **unhappy** / about / all / the noise.
ARC-Sg-Ug	The waitress, / who / sat / near / *the girl*, / unsurprisingly / **were** / **unhappy** / about / all / the noise.
ARC-Pl-Gr	The waitress, / who / sat / near / *the girls*, / unsurprisingly / **was** / **unhappy** / about / all / the noise.
ARC-Pl-Ug	The waitress, / who / sat / near / *the girls*, / unsurprisingly / **were** / **unhappy** / about / all / the noise.
RRC-Sg-Gr	The waitress / who / sat / near / *the girl* / unsurprisingly / **was** / **unhappy** / about / all / the noise.
RRC-Sg-Ug	The waitress / who / sat / near / *the girl* / unsurprisingly / **were** / **unhappy** / about / all / the noise.
RRC-Pl-Gr	The waitress / who / sat / near / *the girls* / unsurprisingly / **was** / **unhappy** / about / all / the noise.
RRC-Pl-Ug	The waitress / who / sat / near / *the girls* / unsurprisingly / **were** / **unhappy** / about / all / the noise.

Note. Distractor noun italicized. Regions of interest are bold-faced. The "/" sign indicates regions.

Abbreviations: ARC, appositive relative clause; Gr, grammatical; Pl, plural distractor noun; RRC, restrictive relative clause; Sg, singular distractor noun; Ug, ungrammatical.

target subject was always modified with a subject-extracted RC, which contained the distractor noun, NP2. The distractor noun was either a singular noun or a plural noun (e.g., *the girl(s)*). The RC was either an RRC or an ARC, and the ARC conditions were marked with commas before and after the RC boundary. There was always an intervening adverb between the distractor noun and the main verb (e.g., *unsurprisingly*). The number of regions for the target trials ranged from 11 to 13. We also included 24 filler sentences, which were all grammatical sentences.

4.1.3. Procedure

We used a Latin-square design to assign the 48 main trial items into eight lists. Each participant read a total of 72 sentences (including filler sentences), with half of the sentences being grammatical and the other half ungrammatical. The presentation of the trials was randomized for each participant.

The experiment was conducted on IbexFarm (http://spellout.net/ibexfarm), an online experiment platform. Participants read sentences in a self-paced phrase-by-phrase moving window paradigm. Each sentence was presented individually on each participant's screen. Words were initially masked by dashes, and each word or phrase appeared as the participant pressed the space bar. The sentence was presented in a noncumulative fashion, where the previous word was masked again by dashes as the participant proceeded to the next region. No line break was included in any of the stimuli, and all of them were presented in a single line. Participants were instructed to read the sentences as naturally as possible at their regular reading pace. A "yes/no" comprehension task that asked about the content of the sentence appeared after each sentence (e.g., "Was the waitress unhappy about all the noise?"). Half of the questions had "yes" correct responses and the other half "no." The participants were not given any feedback on their responses to the task. Participants had 10 practice trials before the main experiment.

4.1.4. Analysis

Prior to data analysis, we excluded results from 10 participants, whose accuracy on the comprehension question task for both target and filler items was below 80%. For the reading time analysis for the self-paced reading task, trials with incorrect comprehension question

Table 2
Mean comprehension question accuracy in Experiment 1

	Gr-Pl	Gr-Sg	Ug-Pl	Ug-Sg
ARC	0.935 (±0.003)	0.925 (±0.003)	0.910 (±0.003)	0.937 (±0.003)
RRC	0.927 (±0.003)	0.934 (±0.003)	0.913 (±0.003)	0.922 (±0.003)

Note. Standard errors of the grand mean are in parentheses.

Abbreviations: ARC, appositive relative clause; Gr, grammatical; Pl, plural distractor noun; RRC, restrictive relative clause; Sg, singular distractor noun; Ug, ungrammatical.

responses were removed for the main analysis (7.47% of the data). We removed reading time data points that were beyond 3SD of the mean by condition and by region (additional 1.5% removal of the data). Regions of interest included (a) the main verb position (*was/were*) (region 7) that agreed with the target subject (*the waitress*) and (b) the next spillover region (*unhappy*) (region 8).

We analyzed the log-transformed reading time data by conducting linear mixed-effects regression models, using the `lmerTest` package (Bates, Mächler, Bolker, & Walker, 2015) implemented in R (R Core Team, 2022). We ran two different models. In the first model, fixed effects included grammaticality (grammatical vs. ungrammatical), distractor number (singular vs. plural), clause type (ARC vs. RRC), and their interactions. In addition to this regular model, since our main interest was whether RRCs and ARCs demonstrate different agreement attraction profiles, we also conducted a second model using a nested contrast to estimate the number agreement attraction effect under each level of the clause type. Following the nested contrast approach in Nicenboim, Schad, and Vasishth (2023), we set up a customized contrast coding that estimated the two-way interaction between distractor number and grammaticality as a nested effect for ARC and RRC separately.[5] For both models, the log-transformed reading time of the immediately preceding region was also included as a fixed effect, to take into consideration of the spillover effect from regions prior to the current region.

All models were initially fit with maximal random effects structure with by-subject and by-item random intercepts and by-subject and by-item random slopes for all fixed-effect predictors (Barr, Levy, Scheepers, & Tily, 2013). Random effects structure was simplified when the models failed to converge. We used the `buildmer` package (Voeten, 2022) in R to find the maximal model that would converge. A fixed effect was considered significant when the absolute *t*-value associated with the effect was equal or exceeded 2 (Gelman & Hill, 2006). The tables for the summary of the statistical analyses will present the fixed effects and random effects included in the converged maximal model.

4.2. Results

4.2.1. Comprehension accuracy

The mean comprehension question accuracy for each condition is presented in Table 2. The comprehension accuracy for all conditions was generally close to ceiling.

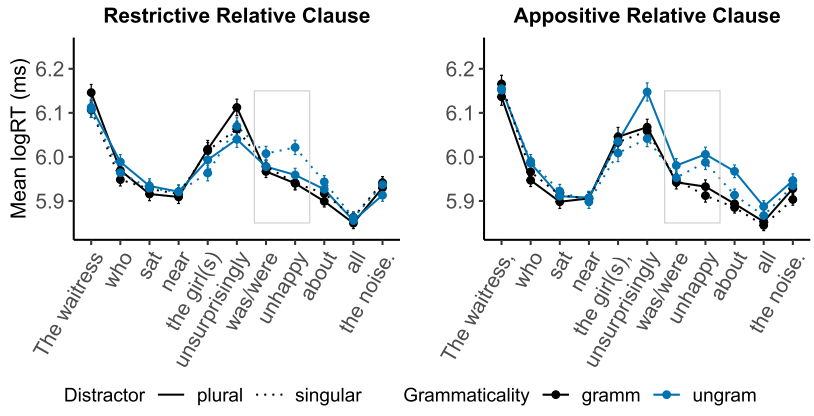


Fig. 2. Mean log reading times in Experiment 1. Error bars represent standard errors.

Table 3
Summary of the regular mixed effects model of the reading times in Experiment 1

	Regions					
	critical			spillover		
Fixed effects	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*
(Intercept)	4.602	0.065	71.22	3.827	0.074	51.85
logRT of the previous region	0.225	0.010	22.08	0.357	0.012	29.83
Clause	**0.030**	**0.008**	**3.96**	−0.002	0.009	−0.26
Distractor	0.010	0.008	1.37	0.007	0.007	0.94
Grammaticality	**0.026**	**0.008**	**3.22**	**0.058**	**0.009**	**6.41**
Clause:Distractor	0.018	0.015	1.19	**0.031**	**0.015**	**2.15**
Clause:Grammaticality	0.018	0.015	1.21	−0.019	0.015	−1.28
Distractor:Grammaticality	−0.012	0.015	−0.81	**0.029**	**0.015**	**2.00**
Clause:Distractor:Grammaticality	0.013	0.030	0.42	0.024	0.029	0.84

Random intercepts and slopes (slope in parenthesis)	Var.	*SD*	Var.	*SD*
Subject	0.033	0.182	0.029	0.170
Subject (Grammaticality)	0.001	0.032	0.003	0.055
Subject (Clause)	—	—	0.002	0.047
Item	0.000	0.012	0.001	0.034

Note. The effects of the random slopes that were not included in the final converged model are not presented in the table.

4.2.2. Self-paced reading task

Fig. 2 presents the mean log reading times for each region, with regions of interest highlighted in gray boxes.

Table 3 shows the statistical results with the three fixed effects and their interactions included in the model. We found a main effect of grammaticality (ungrammatical > grammatical) in both the critical ($\hat{\beta} = 0.026$, $se = 0.008$, $t = 3.22$) and spillover ($\hat{\beta} = 0.058$, $se =$

0.009, $t = 6.41$) regions. The RRC condition took longer than the ARC condition at the critical region ($\hat{\beta} = 0.030$, $se = 0.008$, $t = 3.96$). More relevant for our purpose, at the spillover region, there was a significant two-way interaction of clause and distractor ($\hat{\beta} = 0.031$, $se = 0.015$, $t = 2.15$) and a significant two-way interaction of distractor and grammaticality ($\hat{\beta} = 0.029$, $se = 0.015$, $t = 2.0$); yet, the three-way interaction did not reach significance in either of the critical ($\hat{\beta} = 0.013$, $se = 0.030$, $t = 0.42$) or the spillover region ($\hat{\beta} = 0.024$, $se = 0.029$, $t = 0.84$). Since we observed a two-way interaction of distractor and grammaticality in the spillover region, and the agreement attraction effect has previously been shown to be more robust in ungrammatical sentences (e.g., Jäger, Mertzen, Van Dyke, & Vasishth, 2020; Wagers et al., 2009; Yadav, Paape, Smith, Dillon, & Vasishth, 2022; Yadav, Smith, Reich, & Vasishth, 2023), we conducted a post hoc analysis on the spillover region by separately analyzing the grammatical and ungrammatical conditions. We found a two-way interaction between clause and distractor only in the ungrammatical conditions ($\hat{\beta} = 0.043$, $se = 0.022$, $t = 2.0$) but not in the grammatical condition ($\hat{\beta} = 0.021$, $se = 0.020$, $t = 1.06$). The two-way interaction of clause and distractor in the ungrammatical conditions was driven by the fact in the RRC ungrammatical conditions, the plural distractor condition was faster than the singular distractor condition ($\hat{\beta} = 0.043$, $se = 0.015$, $t = 2.81$), indicating an agreement attraction effect; but there was no difference between the plural versus singular distractor conditions for ARCs ($\hat{\beta} = 0.001$, $se = 0.015$, $t = 0.05$).

Table 4 presents the results of the statistical analysis with the nested model, that is, the effects of the distractor, grammaticality, and their interactions nested under each level of the predictor clause type. The analysis shows that at the critical region, there was a main effect of clause type ($\hat{\beta} = 0.11$, $se = 0.005$, $t = 2.30$) such that RRC conditions were read slower than ARC conditions; at the spillover region, ARC conditions were read slower than RRC conditions ($\hat{\beta} = -0.018$, $se = 0.005$, $t = -3.35$). Since this is not our main effect of interest, we will not pursue it further. Both RRCs and ARCs showed a grammaticality effect. For the RRCs, this effect appeared on both the critical ($\hat{\beta} = 0.018$, $se = 0.005$, $t = 3.32$) and the spillover region ($\hat{\beta} = 0.024$, $se = 0.005$, $t = 4.73$). For the ARCs, this effect appeared on the spillover region ($\hat{\beta} = 0.067$, $se = 0.012$, $t = 5.52$), but not on the critical word ($\hat{\beta} = 0.017$, $se = 0.011$, $t = 1.60$). Most relevant for the current purpose, at the spillover region, there was an interaction of distractor and grammaticality within the RRC condition ($\hat{\beta} = 0.010$, $se = 0.005$, $t = 2.01$), indicating the agreement attraction effect. This interaction on the RRC conditions was driven by longer reading times in the singular-distractor condition compared to the plural-distractor condition for the ungrammatical trials ($\hat{\beta} = 0.043$, $se = 0.015$, $t = 2.81$), but not for the grammatical ones ($\hat{\beta} = 0.000$, $se = 0.014$, $t = 0.024$). Crucially, for the ARC conditions, no interaction between distractor and grammaticality was found on either the critical region ($\hat{\beta} = -0.009$, $se = 0.011$, $t = -0.88$) or the spillover region ($\hat{\beta} = 0.008$, $se = 0.010$, $t = 0.80$).

Fig. 3 shows the interference effect of each clause type at the critical and spillover regions. The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition, and, therefore, a negative value is indicative of the agreement attraction effect (as in Dillon et al., 2013 and Jäger et al., 2017). There was a significant attraction effect only at the spillover region in the ungrammat-

Table 4
Summary of the nested contrast model of the reading times in Experiment 1

	Regions					
	critical			spillover		
Fixed effects	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*
(Intercept)	4.575	0.065	70.30	3.839	0.074	51.92
logRT of the previous region	0.230	0.010	21.91	0.358	0.012	29.89
Clause	**0.11**	**0.005**	**2.30**	**−0.018**	**0.005**	**−3.35**
Distractor in ARC	−0.004	0.008	−0.54	−0.000	0.007	−0.06
Grammaticality in ARC	0.017	0.011	1.60	**0.067**	**0.012**	**5.52**
Distractor in RRC	0.010	0.005	1.84	**0.011**	**0.005**	**2.20**
Grammaticality in RRC	**0.018**	**0.005**	**3.32**	**0.024**	**0.005**	**4.73**
Distractor:Grammaticality in ARC	−0.009	0.011	−0.88	0.008	0.010	0.80
Distractor:Grammaticality in RRC	−0.002	0.005	−0.29	**0.010**	**0.005**	**2.01**
Random intercepts and slopes (slope in parenthesis)		Var.	*SD*	Var.	*SD*	
Subject		0.027	0.165	0.029	0.171	
Subject (logRT of previous region)		0.000	0.019	—	—	
Subject (Clause)		—	—	0.001	0.030	
Subject (Grammaticality in ARC)		—	—	0.004	0.065	
Item		0.000	0.012	0.001	0.034	

Note. The effects of the random slopes that were not included in the final converged model are not presented in the table.

ical RRC conditions. The interference effect for each participant can be found in Appendix A (Fig. A1).

4.3. Discussion

Taken together, the results from Experiment 1 found an agreement attraction effect in RRCs but not in ARCs, in particular on the spillover region. This is most clearly shown in the mixed effects model that nested the agreement attraction effect under each clause type. In this analysis, the interaction between distractor and grammaticality, which was used as an index for the attraction effect, was only found for RRCs. There is a notable caveat that we did not find a three-way interaction between clause type, distractor, and grammaticality in the regular full mixed effects model. We will return to more discussion about this caveat in Section 7.4. But it is worth noting that there was a significant distractor by grammaticality interaction at the spillover region in the regular full model. This interaction was driven by the fact that only in the ungrammatical sentences but not in the grammatical ones there was a further interaction between clause type and distractor, such that the agreement attraction effect was only observed for the RRCs but not for the ARCs. The lack of three-way interaction may be driven by the fact that the number interference effect in grammatical sentences is noisier and more variable, making it difficult to detect a three-way interaction.

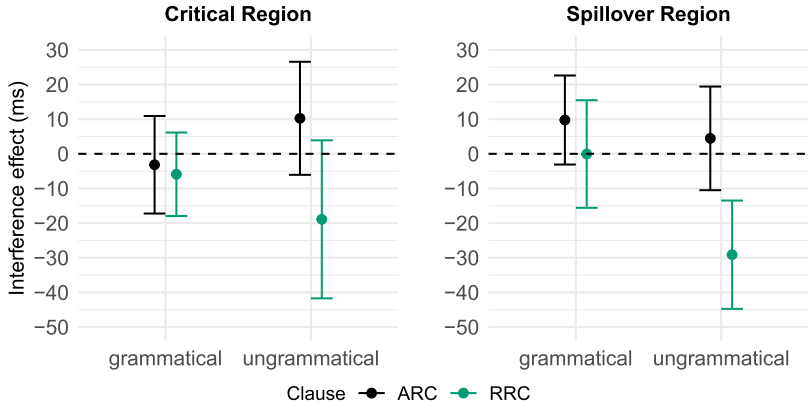


Fig. 3. Interference effect in Experiment 1. The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition. Error bars indicate 95% CI by participants.

The agreement attraction effect on the RRC conditions replicated the finding in Parker and An (2018) (Experiment 1), where there was a standard number agreement attraction effect in the presence of a retrieval cue that matched the number feature of the distractor noun. The results are in line with a large body of existing findings on the cue-based retrieval mechanism in long-distance dependency resolution, evidenced by the interference effect (e.g., Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & McElree, 2011). Crucially, however, no significant agreement attraction effect was found in the ARC condition. This is in line with the finding in Ng and Husband (2017) and McInnerney and Atkinson (2020), where the intervening distractor noun within a subordinate discourse unit did not interfere with the target noun during the retrieval process. The current findings are also in line with earlier work that showed a separation between ARCs and RRCs within the context of syntactic complexity (e.g., Dillon et al., 2014; Kroll and Wagers, 2019).

There are two ways to interpret the observed difference between RRCs and ARCs in Experiment 1. One interpretation is that the main and subordinate discourse units can be separated during memory encoding and retrieval by virtue of their distinct linguistic status. As discussed earlier, the linguistics literature has long recognized that different types of discourse units can have distinct representational statuses, with proposals suggesting that discourse units are stacked in the "attentional space" (Grosz and Sidner, 1986) one at a time in such a way that each unit can be independent of one another, or different discourse units could contribute to distinct semantic dimensions (Potts, 2005), or the subordinate discourse units are similar to speech acts that carry independent illocutionary functions distinct from the main proposition (Frazier et al., 2018; Koev, 2022). A way to implement the linguistic distinction between main and subordinate discourse status in a processing model is to assume that the parser encodes the discourse status information as features of the relevant linguistic input (e.g., [+MAIN] or [+SUBORDINATE]). These binary and static features can be used as retrieval cues to guide

memory retrieval. As a result, the parser can make a distinction in real time between linguistic units based on their discourse status.

Another interpretation of the results is based on the notion of *active discourse question*. As discussed in Section 3.1, from the perspective that discourse structure is incrementally constructed, subordinate discourse information can raise active discourse questions but such questions become inactive once the subordinate discourse information is closed off. Discourse structure moves forward based on which questions are raised and remain active and which questions have been resolved and can be removed. We can make a further assumption that working memory retrieval, or for the purpose of subject-verb agreement dependency at least, is sensitive to the information currently in the domain of active questions. With this approach, we can explain the results by the distractor noun inside the subordinate ARC becoming no longer accessible to memory retrieval after the closure of the ARC, at the point of retrieval. This made the target noun, which was inside the main discourse unit, become the only candidate for memory retrieval, leading to the absence of an agreement attraction effect in the ARC condition.

Both of the proposals above are consistent with the findings from Experiment 1, in which memory retrieval for an agreement-controller noun takes place after the closure of the ARC clause. However, the two proposals rely on distinct theoretical assumptions. The first proposal, which we label the *static division hypothesis*, assumes that the working memory organization of the linguistic material is sensitive to the static properties of the discourse structure such as the main versus subordinate characteristics. The second proposal, which we label the *active question hypothesis*, assumes that the working memory organization incrementally tracks the relevant questions/issues at any given moment of an unfolding sentence. To distinguish these two proposals, in the next experiment, we modify the design such that memory retrieval of a target takes place prior to the closure of the discourse question instantiated by the subordinate ARC.

5. Experiment 2

In Experiment 2, we locate the memory retrieval site prior to the subordinate ARC being closed off. The design of Experiments 1 and 2 are compared in (7).

(7) Schematization of experimental design (Experiments 1–2)
   a. Experiment 1
      Target(,) [who… ***Distractor***…](,) **RetrievalSite**…
   b. Experiment 2
      Distractor(,) [who… **Target RetrievalSite**…](,)…

For the baseline control RRC conditions, we expect to observe a standard agreement attraction effect in Experiment 2, replicating the basic attraction effect in Experiment 1. As for the critical ARC conditions, the two competing hypotheses make contrasting predictions. The *static division hypothesis* predicts that the ARC conditions in Experiment 2 would demonstrate a lack of attraction effect, similar to Experiment 1. This is because the correct retrieval target

Table 5
Sample set of experimental items for Experiment 2

Condition	Sentence
ARC-Sg-Gr	The / *musician*, / who / the / reviewer / **praises** / **so** / highly, / will / probably / win / a / Grammy.
ARC-Sg-Ug	The / *musician*, / who / the / reviewer / **praise** / **so** / highly, / will / probably / win / a / Grammy.
ARC-Pl-Gr	The / *musicians*, / who / the / reviewer / **praises** / **so** / highly, / will / probably / win / a / Grammy.
ARC-Pl-Ug	The / *musicians*, / who / the / reviewer/ **praise** / **so** / highly, / will / probably / win / a / Grammy.
RRC-Sg-Gr	The / *musician* / who / the / reviewer / **praises** / **so** / highly / will / probably / win / a / Grammy.
RRC-Sg-Ug	The / *musician* / who / the / reviewer / **praise** / **so** / highly / will / probably / win / a / Grammy.
RRC-Pl-Gr	The / *musicians* / who / the / reviewer / **praises** / **so** / highly / will / probably / win / a / Grammy.
RRC-Pl-Ug	The / *musicians* / who / the / reviewer / **praise** / **so** / highly / will / probably / win / a / Grammy.

Note. Distractor noun italicized. Regions of interest are bold-faced. The "/" sign indicates regions.

Abbreviations: ARC, appositive relative clause; Gr, grammatical; Pl, plural distractor noun; RRC, restrictive relative clause; Sg, singular distractor noun; Ug, ungrammatical.

NP and the distractor NP in the ARC condition are located in different discourse units in both Experiments 1 and 2. In Experiment 1, the target NP is located in the main unit, while the distractor NP is in the subordinate unit; in Experiment 2, it is the other way around. If the parser uses discourse status information [+MAIN] or [+SUBORDINATE] to guide retrieval, the target and distractor NP should be sufficiently distinguished from each other, reducing the probability of an interference effect. On the other hand, the *active question hypothesis* would make a different prediction. Under this hypothesis, we would expect to see an attraction effect in Experiment 2, different from Experiment 1. This is because the retrieval site in Experiment 2 is located prior to the closure of the ARC (it is inside the ARC). This means that at the point where memory retrieval of an agreement controller is initiated, the discourse questions associated with the main and the subordinate discourse units are both active, and accordingly, linguistic content in the scope of these questions is accessible for retrieval. This leads to the possibility of misretrieving the distractor, resulting in a retrieval interference effect.

5.1. Methods

5.1.1. Subjects

Experiment 2 recruited a total of 120 American English speakers residing in the United States (aged 18 and above) through the Prolific platform (https://www.prolific.co/).[6] Participation recruitment was conducted with the IRB approval of the local institution. Two participants, indicating English was not their first language, were excluded, and an additional two participants who did not complete the experiment were removed, resulting in a total of 116 participants (mean age = 31.15; range: 18–50). The experiment, lasting approximately 20 min, provided participants with 3.50–4.00 USD in compensation.

5.1.2. Materials

Material for Experiment 2 is presented in Table 5. The material consisted of 48 sets of items with 8 conditions in a fully crossed 2 x 2 x 2 design with grammaticality (grammatical

vs. ungrammatical), distractor number (singular vs. plural), and clause type (ARC vs. RRC) as factors. We adopted the RRC material from Experiment 2 in Wagers et al. (2009). All the target sentences had the structure of *NP1(,) who NP2 VERB..., VERB...* (e.g., "The musician(s)(,) who the reviewer praise(e) so highly, will…"). NP1 was the distractor noun and appeared at the matrix subject position, which was both structurally and linearly distant from the target verb. It was varied by its number feature, either a singular or plural noun (e.g., *the musician(s)*). The sentence always had an object-extracted RC structure, where NP1 was the extracted object, and NP2 was the subject in the embedded clause. NP2 was the target noun, which formed a number-agreement dependency relation with the target verb that appeared one region after the target noun and inside the RC. NP2 was always a singular noun (e.g., *the reviewer*), and the target verb either agreed with the target noun (e.g., *praises*) or not (e.g., *praise*), determining the grammaticality of the entire sentence. ARC structures were marked with commas on the RCs. There were 12–20 regions for the target trial. We also included 48 grammatical filler sentences, which were taken from the filler items in Wagers et al. (2009).

5.1.3. Procedure

The 48 sets of target items were pseudo-randomly assigned to eight lists under a Latin-square design. Experiment 2 also used a self-paced reading task, conducted on IbexFarm. The experiment was done in a noncumulative word-by-word moving window fashion. There was no linebreak in any of the stimuli. Similar to Experiment 1, we included a "yes/no" comprehension question task that targeted the content of the sentence the participants just saw (e.g., "Will the musician(s) likely win an award?"). Half of the questions targeted "yes" and the other half "no" as a response. Each participant read 96 sentences in total, with two-thirds of the sentences being grammatical. No feedback on their responses was given. There were 10 practice trials before the main experiment.

5.1.4. Analysis

Four participants whose comprehension question accuracy was below 80% were excluded, leaving us with 112 people for the analysis. Trials with incorrect responses in the comprehension question task were removed (5.88% removal). For the self-paced reading task data analysis, we used the same reading time threshold as in Experiment 1 (3SD reading time cutoff by condition and by region), removing an additional 1.53% of the data. There were two critical regions for the main analysis: (a) the target verb region (e.g., *praise(s)*) (region 6) and (b) the spillover region (*so*) (region 7). As in Experiment 1, two regression models were constructed. The first model was a regular mixed effects model with the three fixed-effect predictors and their interactions; the second model used nested contrast to examine the interaction of distractor and grammaticality within each clause type. The tables for the summary of the statistical analyses will present the fixed effects and random effects included in the maximal model.

Table 6
Mean comprehension question accuracy in Experiment 2

	Gr-Pl	Gr-Sg	Ug-Pl	Ug-Sg
ARC	0.944 (±0.944)	0.951 (±0.001)	0.952 (±0.001)	0.928 (±0.002)
RRC	0.958 (±0.001)	0.933 (±0.002)	0.927 (±0.002)	0.937 (±0.002)

Note. Standard errors of the grand mean are in parentheses.

Abbreviations: ARC, appositive relative clause; Gr, grammatical; Pl, plural distractor noun; RRC, restrictive relative clause; Sg, singular distractor noun; Ug, ungrammatical.

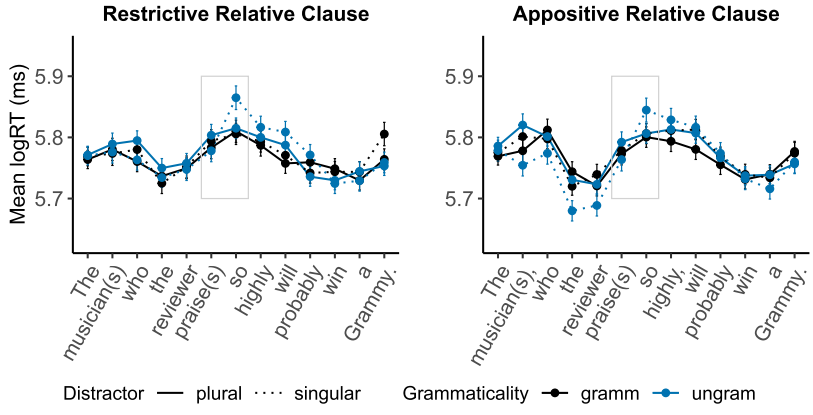


Fig. 4. Mean log reading times in Experiment 2. Error bars represent standard errors.

5.2. Results

5.2.1. Comprehension accuracy

Mean comprehension question accuracy for each condition is presented in Table 6. The comprehension accuracy was high across all conditions.

5.2.2. Self-paced reading task

Fig. 4 shows the mean log reading times in Experiment 2. Table 7 shows the statistical results of the reading times with the model including the three fixed effects and their interactions. No effects were found on the critical region. In the spillover region, there was an effect of distractor and an effect of grammaticality. There was no three-way interaction, but there was a significant two-way interaction effect between distractor and grammaticality ($\hat{\beta} = 0.054$, $se = 0.015$, $t = 3.54$). As in Experiment 1, a post hoc analysis separating grammatical and ungrammatical conditions was conducted. We found no two-way interaction between clause and distractor in either the ungrammatical conditions ($\hat{\beta} = 0.002$, $se = 0.023$, $t = 0.09$); or the grammatical conditions ($\hat{\beta} = -0.002$, $se = 0.020$, $t = -0.12$). Crucially, however, for the ungrammatical conditions (but not for the grammatical ones), there was a main effect of distractor ($\hat{\beta} = 0.055$, $se = 0.011$, $t = 4.88$), indicating an agreement attraction effect for both ARC and RRC clause types, that is, for both clause types, the ungrammat-

Table 7
Summary of the regular mixed effects model for the reading times in Experiment 2

	Regions					
	critical			spillover		
Fixed effects	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*
(Intercept)	3.198	0.073	43.82	3.296	0.068	48.70
logRT of the previous region	0.450	0.012	36.72	0.435	0.011	38.66
Clause	−0.001	0.008	−0.13	0.006	0.008	0.76
Distractor	−0.001	0.008	−0.16	**0.028**	**0.008**	**3.67**
Grammaticality	0.005	0.008	0.68	**0.031**	**0.008**	**4.08**
Clause:Distractor	0.004	0.015	0.25	0.000	0.015	0.03
Clause:Grammaticality	−0.006	0.015	−0.38	0.015	0.015	1.01
Distractor:Grammaticality	−0.014	0.015	−0.94	**0.054**	**0.015**	**3.54**
Clause:Distractor:Grammaticality	0.014	0.030	0.45	0.005	0.030	0.16
Random intercepts	Var.	*SD*		Var.	*SD*	
Subject	0.037	0.193		0.031	0.176	
Item	0.001	0.024		0.001	0.031	

Table 8
Summary of the nested contrast model of the reading times in Experiment 2

	Regions					
	critical			spillover		
Fixed effects	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*
(Intercept)	3.200	0.073	43.83	3.302	0.068	48.76
logRT of the previous region	0.450	0.012	36.72	0.435	0.011	38.66
Clause	−0.002	0.005	−0.54	−0.003	0.005	−0.63
Distractor in ARC	−0.007	0.008	−0.90	**0.027**	**0.008**	**3.52**
Grammaticality in ARC	0.008	0.011	0.75	**0.023**	**0.011**	**2.17**
Distractor in RRC	0.000	0.005	0.07	**0.014**	**0.005**	**2.61**
Grammaticality in RRC	0.001	0.005	0.21	**0.019**	**0.005**	**3.59**
Distractor:Grammaticality in ARC	−0.011	0.011	−0.98	**0.026**	**0.011**	**2.40**
Distractor:Grammaticality in RRC	−0.002	0.005	−0.35	**0.014**	**0.005**	**2.61**
Random intercepts	Var.	*SD*		Var.	*SD*	
Subject	0.037	0.193		0.031	0.176	
Item	0.001	0.024		0.001	0.031	

ical singular-distractor condition was read slower than the ungrammatical plural-distractor condition.

Table 8 reports the statistical analysis of the reading times with the effects of distractor and grammaticality nested under clause type. The analysis found no statistically significant effects in the critical region. In the spillover region, there was a significant interaction of distractor

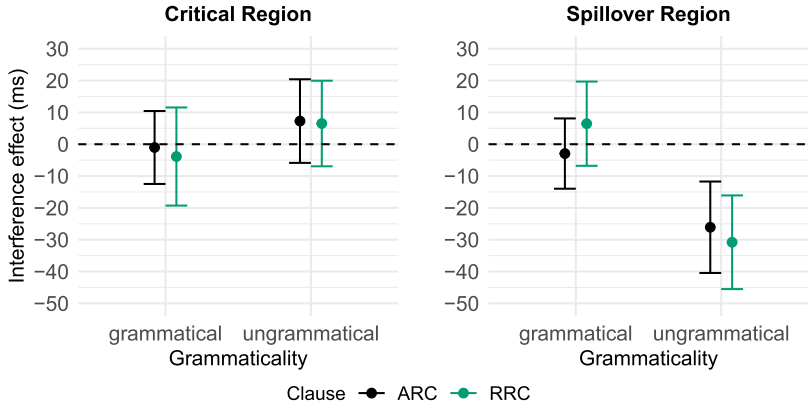


Fig. 5. Interference effect in Experiment 2. The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition. Error bars indicate 95% CI by participants.

and grammaticality both in the ARC condition ($\hat{\beta} = 0.026$, $se = 0.011$, $t = 2.40$) and in the RRC condition ($\hat{\beta} = 0.014$, $se = 0.005$, $t = 2.61$).

Fig. 5 illustrates the interference effect in each clause type at the critical and spillover regions. There was an agreement attraction effect in the ungrammatical condition at the spillover region for both clause types. The interference effect at the individual participant level can be found in Appendix A (Fig. A2).

5.3. Discussion

Experiment 2 found the attraction effect in both clause types. The results were consistent across different models we conducted. We found the agreement attraction effect in the baseline RRC conditions, similar to the RRC conditions in Experiment 1. There was an attraction effect in the RRC conditions even when the distractor noun was further away from the retrieval verb than the target noun, replicating earlier findings using similar constructions (e.g., Wagers et al., 2009). More importantly, different from Experiment 1, the number agreement attraction effect was also observed in the ARC condition. In the current experiment, the distractor inside the main discourse unit interfered with the target inside the subordinate discourse unit.

We return to the question that motivated us to conduct Experiment 2: whether the absence of an interference effect in the ARC condition in Experiment 1 was due to the linguistic units hosting the target and the distractor NPs having distinct discourse status (either [+MAIN] or [+SUBORDINATE]), or the memory retrieval being sensitive to the active state of discourse questions. The current findings lend support to the latter, supporting the active discourse question hypothesis. At the time of memory retrieval, the discourse questions associated with the linguistic units containing the target and the distractor NPs are all active, allowing the distractor to be accessible for memory retrieval and consequently resulting in interference with the target.

There is another remaining possibility, however. Let us consider the design of Experiments 1 and 2 again, as schematized in (7). The above-mentioned conclusion was based on the assumption that the two experiments differ regarding which discourse question(s) are active by the time memory retrieval of the agreement controller is initiated. However, there is another difference between the two experiments. In the design of Experiment 2 (see (7)), the distractor NP is the matrix subject NP ("The musician(s), who the reviewer praise(s)…"), and it signals an upcoming matrix verb. While the parser processes the ARC, the subject NP of the matrix clause (*the musician(s)*) is yet to be integrated with a verb. It is possible that the parser allocates additional resources to maintain the subject NP since it is part of an incomplete syntactic dependency, hence keeping the distractor NP at a relatively high activation level. The presence of the number agreement attraction effect in the ARC condition in Experiment 2, but not in Experiment 1, could be due to the fact that the distractor NP in Experiment 2 had a higher activation level than the distractor NP in Experiment 1, making it a stronger competitor to the target NP in Experiment 2. To rule out this alternative possibility, in Experiment 3 below, we changed the syntactic position of the distractor NP to be an object NP of the matrix clause. In doing so, the distractor NP marks the end of an SVO structure instead of signaling an open dependency.

6. Experiment 3

We modified the material in Experiment 2 such that the distractor NP is now the object NP in the matrix clause. The core SVO argument structure at the matrix clause level is completed by the time the RC structure appears. A schematization of the manipulation is illustrated in (8):

(8) Schematization of experimental design (Experiments 2–3)
   a. Experiment 2
      Distractor(,) [who… **Target RetrievalSite**…](,)…
   b. Experiment 3
      Name… ***Distractor***(,) [who **Target RetrievalSite**…].

In Experiment 3, similar to Experiment 2, the discourse questions associated with the target and the distractor NPs are both active when the memory retrieval of the agreement controller takes place. But different from Experiment 2, in Experiment 3, the distractor NP is an object NP and does not receive an additional activation boost resulting from the active maintenance of an incomplete dependency. If we still observe an agreement attraction effect in the ARC conditions in Experiment 3, this would complement Experiment 2 to provide evidence for the active question-based constraint on memory retrieval.

Table 9
Sample set of experimental items for Experiment 3

Condition	Sentence
ARC-Sg-Gr	Alicia / met / the / *musician*, / who / the / reviewer / **praises** / **so** / highly.
ARC-Sg-Ug	Alicia / met / the / *musician*, / who / the / reviewer/ **praise** / **so** / highly.
ARC-Pl-Gr	Alicia / met / the / *musicians*, / who / the / reviewer / **praises** / **so** / highly.
ARC-Pl-Ug	Alicia / met / the / *musicians*, / who / the / reviewer / **praise** / **so** / highly.
RRC-Sg-Gr	Alicia / met / the / *musician* / who / the / reviewer / **praises** / **so** / highly.
RRC-Sg-Ug	Alicia / met / the / *musician* / who / the / reviewer / **praise** / **so** / highly.
RRC-Pl-Gr	Alicia / met / the / *musicians* / who / the / reviewer / **praises** / **so** / highly.
RRC-Pl-Ug	Alicia / met / the / *musicians* / who / the / reviewer / **praise** / **so** / highly.

Note. Distractor noun italicized. Regions of interest are bold-faced. The "/" sign indicates regions.

Abbreviations: ARC, appositive relative clause; Gr, grammatical; Pl, plural distractor noun; RRC, restrictive relative clause; Sg, singular distractor noun; Ug, ungrammatical.

6.1. Methods

6.1.1. Subjects

A total of 120 native speakers of American English over the age of 18 residing in the United States were recruited via Prolific (https://www.prolific.co/).[7] The study was conducted with IRB approval from the local institution. The duration of the experiment was about 20 min, and participants were paid 3.50–4.00 USD in compensation. Four participants whose self-reported first language was not English were removed. One participant did not complete the experiment. This left us with 115 participants (mean age = 33.37; range: 18–50).

6.1.2. Materials

Table 9 shows an example of material used for Experiment 3. We had 48 sets of 8 items, with 8 conditions, varied by grammaticality (grammatical vs. ungrammatical), distractor number (singular vs. plural), and clause type (ARC vs. RRC). We used the same material in Experiment 2 but modified it in a way that the structure had an SVO structure, followed by an ARC or an RRC. The target material had the structure of *Name VERB NP1(,) who NP2 VERB…* (e.g., "Alicia met the musician(s)(,) who the reviewer praise(s) so highly."), where ARCs were marked with a comma. Forty-eight different names (in the *Name* position) were used for the matrix subject. The number of regions for the target trial ranged from 12 to 14. The same filler sentences in Experiment 2 were used.

6.1.3. Procedure

The 48 target trials, along with 48 filler sentences, which were all grammatical, were distributed across eight lists with a Latin-square design. Each participant was assigned 96 sentences in total. The ratio of grammatical to ungrammatical sentences in the trials was 2:1. There were 13 practice trials before the main trials. We had the same procedure as in Experiment 2, with a self-paced reading task in a noncumulative, moving-window paradigm. No line break was included. Each trial was followed by a comprehension question. Filler trials

Table 10
Mean proportion of selection of the object in Experiment 3

	Gr-Pl	Gr-Sg	Ug-Pl	Ug-Sg
ARC	0.987 (±0.001)	0.976 (±0.002)	0.962 (±0.002)	0.966 (±0.002)
RRC	0.969 (±0.002)	0.960 (±0.002)	0.968 (±0.002)	0.963 (±0.002)

Note. Standard errors of the grand mean are in parentheses.

Abbreviations: ARC, appositive relative clause; Gr, grammatical; Pl, plural distractor noun; RRC, restrictive relative clause; Sg, singular distractor noun; Ug, ungrammatical.

had a "yes/no" comprehension question task that asked a question related to the content of the sentence. Target trials had a forced-choice comprehension question task that probed participants' understanding of the *wh*-pronoun (e.g., "Who does the reviewer praise highly?"). Participants were then asked to select one of the two options, (a) *Alicia* or (b) *the musician(s)*. The order of the choices was randomized.

6.1.4. Analysis

Eleven participants were excluded from the main analysis based on the comprehension question accuracy (below 80%), leaving us with data points from 104 participants for the main analysis. For the reading time result analysis, only the responses that selected the object (e.g., *the musician(s)*) (as opposed to the matrix subject, e.g., *Alicia*) in the question were included for the main data analysis (3.09% data removal). Additionally, reading times beyond 3SD by condition and by region were removed (additional 1.46% data removal). Two regions were identified as regions of interest for the analysis: (a) the target verb region (*praise(s)*) (region 8) and (b) the spillover region (*so*) (region 9). The same analysis methods as in previous experiments were used, including the construction of two mixed-effects models, a regular model and another model with nested effects. The tables for the summary of the statistical analyses will present the fixed effects and random effects included in the converged maximal model.

6.2. Results

6.2.1. Comprehension questions

Table 10 illustrates the mean proportion of selecting the object (rather than the matrix subject) as the antecedent of the *wh*-pronoun. The selection task results indicate a strong tendency among participants to parse the object of the matrix clause (e.g., *the musician(s)*) as the object of the embedded verb (e.g., *praise(s)*).

6.2.2. Self-paced reading task

Mean log reading times for Experiment 3 are presented in Fig. 6. Table 11 presents statistical results using the regular model with the three fixed effects and their interactions. There was a two-way interaction of clause and distractor at the critical region ($\hat{\beta} = 0.034$, $se = 0.015$, $t = 2.26$). In the spillover region, there was a significant two-way interaction of distractor and grammaticality ($\hat{\beta} = 0.041$, $se = 0.015$, $t = 2.82$) along with the main effect of distractor ($\hat{\beta} = 0.021$, $se = 0.008$, $t = 2.54$) and grammaticality ($\hat{\beta} = 0.027$, $se = 0.010$, $t =$

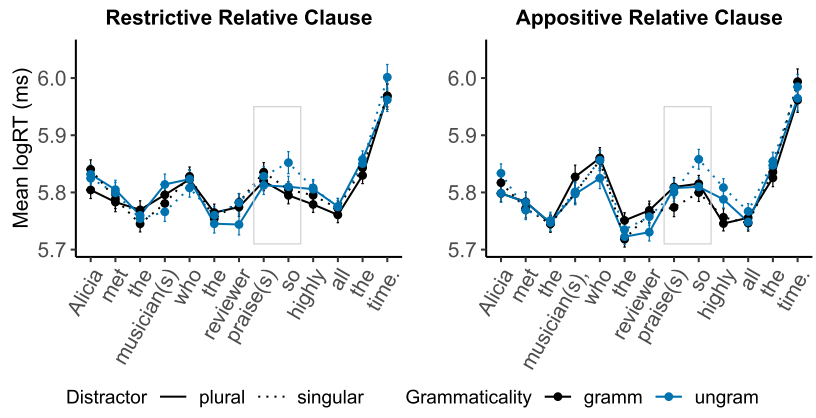


Fig. 6. Mean log reading times in Experiment 3. Error bars indicate the standard error of the mean.

Table 11
Summary of the regular mixed effects model of the reading times in Experiment 3

	Regions					
	critical			spillover		
Fixed effects	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*
(Intercept)	3.837	0.072	52.93	3.422	0.067	51.40
logRT of the previous region	0.341	0.012	28.34	0.411	0.011	37.08
Clause	0.012	0.008	1.53	−0.010	0.008	−1.20
Distractor	−0.009	0.008	−1.12	**0.021**	**0.008**	**2.54**
Grammaticality	0.003	0.008	0.37	**0.027**	**0.010**	**2.85**
Clause:Distractor	**0.034**	**0.015**	**2.26**	−0.001	0.015	−0.09
Clause:Grammaticality	−0.026	0.015	−1.69	0.006	0.015	0.42
Distractor:Grammaticality	0.005	0.015	0.35	**0.041**	**0.015**	**2.82**
Clause:Distractor:Grammaticality	−0.019	0.030	−0.64	−0.017	0.029	−0.59
Random intercepts and slopes (slope in parenthesis)		Var.	*SD*	Var.	*SD*	
Subject		0.039	0.198	0.024	0.156	
Subject (Grammaticality)		—	—	0.004	0.062	
Subject (Distractor)		—	—	0.002	0.043	
Item		0.000	0.020	0.000	0.022	
Item (Clause)		—	—	0.001	0.028	

Note. The effects of the random slopes that were not included in the final converged model are not presented in the table.

2.85). There was no three-way interaction ($\hat{\beta} = -0.017$, $se = 0.029$, $t = -0.59$). To further understand the interaction between grammaticality and distractor in the spillover region, we again conducted a post hoc analysis separating grammatical and ungrammatical conditions at the spillover region. We found a main effect of distractor in the ungrammatical conditions ($\hat{\beta} = 0.041$, $se = 0.011$, $t = 3.56$); but there was no two-way interaction between clause and

Table 12
Summary of the nested contrast model of the reading times in Experiment 3

	Regions					
	critical			spillover		
Fixed effects	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*
(Intercept)	3.841	0.073	52.98	3.310	0.075	44.05
logRT of the previous region	0.341	0.012	28.34	0.409	0.011	36.59
Clause	0.002	0.005	0.41	**−0.011**	**0.005**	**−2.46**
Distractor in ARC	−0.009	0.008	−1.20	**0.024**	**0.007**	**3.28**
Grammaticality in ARC	0.016	0.011	1.45	**0.024**	**0.010**	**2.32**
Distractor in RRC	0.004	0.005	0.81	0.010	0.005	1.95
Grammaticality in RRC	−0.005	0.005	−0.93	**0.015**	**0.007**	**2.33**
Distractor:Grammaticality in ARC	0.007	0.011	0.70	**0.026**	**0.010**	**2.51**
Distractor:Grammaticality in RRC	−0.001	0.005	−0.20	0.008	0.005	1.56
Random intercepts and slopes (slope in parenthesis)		Var.	*SD*	Var.	*SD*	
Subject		0.039	0.198	0.025	0.157	
Subject (Grammaticality in RRC)		—	—	0.002	0.040	
Item		0.000	0.020	0.000	0.022	

Note. The effects of the random slopes that were not included in the final converged model are not presented in the table.

distractor in the ungrammatical trials ($\hat{\beta} = -0.014$, $se = 0.023$, $t = -0.61$). For the grammatical trials, there was neither an effect of distractor ($\hat{\beta} = 0.000$, $se = 0.009$, $t = 0.02$) nor an interaction between clause and distractor ($\hat{\beta} = 0.011$, $se = 0.018$, $t = 0.58$). This result is consistent with the findings in Experiment 2, indicating a standard number agreement attraction effect in both clause types, in particular, in ungrammatical sentences.

Table 12 shows the statistical analysis of reading time results with the model using the nested contrast. No statistically significant effects were found in the critical region. In the spillover region, there was an interaction of distractor and grammaticality in the ARC condition ($\hat{\beta} = 0.026$, $se = 0.010$, $t = 2.51$), indicating an agreement attraction effect. There was a main effect of grammaticality ($\hat{\beta} = 0.015$, $se = 0.007$, $t = 2.33$) and a marginal effect of distractor ($\hat{\beta} = 0.010$, $se = 0.005$, $t = 1.95$) in the RRC condition. However, we did not find an interaction of distractor and grammaticality in the RRC condition ($\hat{\beta} = 0.008$, $se = 0.005$, $t = 1.56$).

The mean interference effects are presented in Fig. 7. Individual participants' interference effect results are presented in Appendix A (Fig. A3).

6.3. Discussion

We again observed the number agreement attraction effect in the ARC conditions in Experiment 3, where the distractor NP was not part of an incomplete dependency. The results of Experiments 2 and 3 together provide evidence for the hypothesis that information within

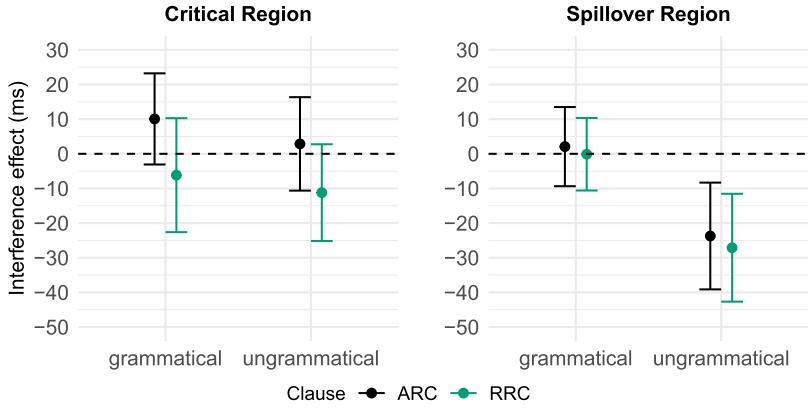


Fig. 7. Interference effect in Experiment 3. The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition. Error bars indicate 95% CI by participants.

main and subordinate discourse units can interfere with each other during memory retrieval when they are both associated with active discourse questions. It is not entirely clear to us why there was not a significant agreement attraction effect with the RRC conditions in Experiment 3. There was a numerical trend of a standard agreement attraction effect in the RRCs, with an effect size of around 30 ms (Fig. 7), which is in line with the effect sizes reported in previous studies (Jäger et al., 2017, for a meta-analysis review). But it is worth noting that some previous work has shown similar results indicating that retrieval interference effects are weaker when the distractor is an object rather than a subject (Parker and An, 2018; Van Dyke & McElree, 2011), possibly due to reduced prominence of object NPs relative to subject ones.

6.4. Between-experiment comparison

In summary, all three experiments found a standard agreement attraction effect for RRC conditions. But for our critical ARC conditions, we found no agreement attraction effect in Experiment 1, but observed a robust attraction effect in Experiments 2 and 3. In this section, we present two additional analyses that directly compare the effect of agreement attraction on the critical ARC conditions across experiments.

First, we analyzed the log-transformed reading time data only from the ARC conditions at the spillover region. We conducted a mixed-effects linear regression model with fixed effects of grammaticality (grammatical vs. ungrammatical), distractor number (singular vs. plural), experiment (Experiments 1–3), and their interactions; word length was also included as a fixed effect to account for the lexical variation across experiments. By-subject and by-item random intercepts were also included. The grammaticality and distractor variables were sum-coded (grammatical = −0.5; ungrammatical = 0.5; plural = −0.5; singular = 0.5). For the experiment variable, since the key observation in the current work comes from the contrast between Experiment 1 vs. Experiments 2 and 3, we used a Helmert effect coding scheme to compare Experiment 1 with Experiments 2 and 3. The Helmert coding also allowed us to

Table 13
Summary of statistical analyses of the reading times across experiments in the ARC condition at the spillover region

	Estimate	Std. Error	*t* value
(Intercept)	5.818	0.019	314.20
Distractor	0.010	0.007	1.45
Grammaticality	**0.041**	**0.007**	**6.25**
Experiment(1 vs. 2–3)	**0.113**	**0.035**	**3.24**
Experiment(2 vs. 3)	**−0.014**	**0.060**	**−0.24**
Distractor:Grammaticality	**0.035**	**0.013**	**2.65**
Distractor:Experiment(1 vs. 2–3)	**−0.029**	**0.014**	**−2.06**
Distractor:Experiment(2 vs. 3)	0.014	0.024	0.59
Grammaticality:Experiment(1 vs. 2–3)	**0.049**	**0.014**	**3.49**
Grammaticality:Experiment(2 vs. 3)	−0.004	0.024	−0.15
Distractor:Grammaticality:Experiment(1 vs. 2–3)	−0.043	0.028	−1.52
Distractor:Grammaticality:Experiment(2 vs. 3)	−0.031	0.048	−0.64

Table 14
Summary of statistical analyses of the reading times across experiments in the ARC ungrammatical conditions at the spillover region

	Estimate	Std. Error	*t* value
(Intercept)	5.835	0.021	280.54
Distractor	**0.026**	**0.010**	**2.65**
Experiment(1 vs. 2–3)	**0.136**	**0.037**	**3.66**
Experiment(2 vs. 3)	−0.016	0.063	−0.25
Distractor:Experiment(1 vs. 2–3)	**−0.050**	**0.021**	**−2.37**
Distractor:Experiment(2 vs. 3)	−0.002	0.036	−0.06

compare Experiment 2 with Experiment 3. The summary of the model is shown in Table 13. Although the variables distractor and grammaticality each had an interaction with the contrast Experiment 1 versus 2–3 (Distractor:Experiment $\hat{\beta} = -0.029$, $se = 0.014$, $t = -2.06$; Grammaticality:Experiment $\hat{\beta} = 0.049$, $se = 0.014$, $t = 3.49$), there was no significant three-way interaction ($\hat{\beta} = -0.043$, $se = 0.028$, $t = -1.52$). It is possible that we did not have sufficient power for a three-way interaction to emerge. We revisit this issue in Section 7.4.

As the agreement attraction effect in the current study was observed only in ungrammatical sentences, we also conducted a second analysis focusing exclusively on the ungrammatical ARC conditions. We constructed a mixed-effects model similar to the one in the previous analysis, but removing the grammaticality variable. As summarized in Table 14, the analysis showed a statistically significant interaction between distractor and experiment, specifically the distractor effect being modulated by the difference between Experiment 1 versus Experiments 2–3 ($\hat{\beta} = -0.050$, $se = 0.021$, $t = -2.37$), driven by the fact that the number agreement attraction effect in the ARC condition in Experiment 1 differed from the effect in Experiments 2–3.

Table 15
Summary of the key experimental design features and results

Exp.	Clause	Discourse status [Target, Distractor]	Active state of discourse questions [Target, Distractor]	Number agreement attraction
1	RRC	[+MAIN, +MAIN]	[+ACTIVE, +ACTIVE]	yes
	ARC	[+**MAIN**, +**SUBORDINATE**]	[+**ACTIVE**, −**ACTIVE**]	**no**
2 and 3	RRC	[+MAIN, +MAIN]	[+ACTIVE, +ACTIVE]	yes
	ARC	[+**SUBORDINATE**, +**MAIN**]	[+**ACTIVE**, +**ACTIVE**]	**yes**

7. General discussion

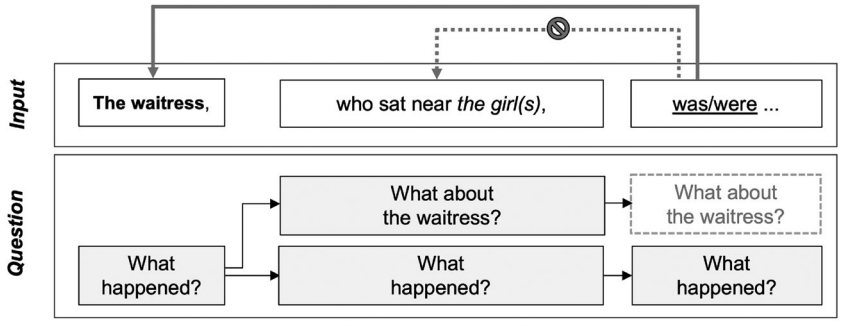
In three self-paced reading experiments, we explored whether discourse structure has an impact on the online construction of subject-verb agreement dependency, in particular the agreement attraction effect. We manipulated discourse structure through the contrast between ARCs and RRCs, where linguistic content encoded by the ARC is a part of the subordinate discourse structure, and the content encoded by the RRC is part of the main discourse structure. Since the subject-verb agreement attraction effect has generally been accounted for by the cue-based memory retrieval model, our findings have implications for the interaction between discourse structure, syntactic structure building, and memory retrieval mechanisms. We also discuss alternative accounts of our data and future directions.

7.1. The effect of active discourse question on agreement attraction

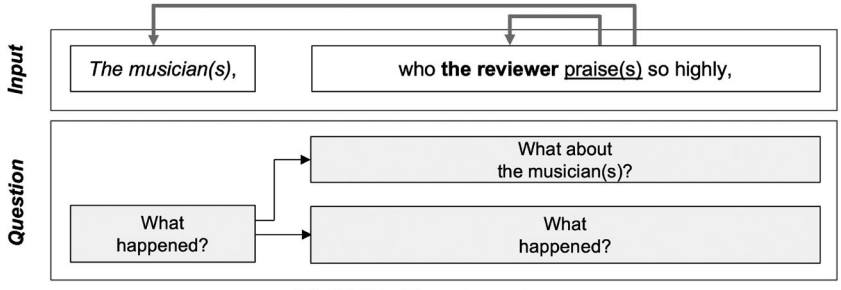
As mentioned earlier, one way to construe the difference between RRCs and ARCs is to establish a static distinction between their discourse status: RRCs encode main discourse information, and ARCs encode subordinate discourse information, at least for the type of stimuli we used in this study. For each experimental condition in Experiments 1–3, we can then classify whether the agreement controller and the potential distractor noun are hosted in a [+MAIN] or [+SUBORDINATE] discourse unit, depending on whether the agreement controller and the distractor are placed in an RRC or ARC. This is summarized in Table 15 under the middle column "Discourse status [Target, Distractor]." If discourse status serves as a cue to guide the memory access of the relevant noun phrase that forms a syntactic dependency with the matrix verb, agreement attraction is expected for the RRC conditions in all three experiments, since the distractor shares the [+MAIN] status with the target NP, giving rise to similarity-based interference. This is indeed what we found, as shown by the summary of the empirical findings in Table 15 under the last column "Number agreement attraction." The static distinction between RRC and ARC, however, fails to capture the findings in the ARC conditions. In all three experiments, for the ARC conditions, the target and distractor NPs had different discourse statuses, making the two NPs featurally distinct and, therefore, predicting no interference/attraction effect for all the ARC conditions. This prediction was not borne out. We did not observe attraction in the ARC condition in Experiment 1, but there was interference in the ARC condition in Experiments 2–3.

To better capture the empirical findings, we use a question-based approach (Jasinskaja, 2016; Riester, 2019) to model how discourse structure is incrementally constructed and how that incremental process exerts an effect on the memory access of representations with different discourse statuses. In particular, we assume that a discourse structure can be represented as a question stack. Different discourse issues/questions arise while a discourse progresses, and these questions stay active until they are resolved and removed from the question stack. In our proposal, representations associated with active discourse questions are more accessible for memory retrieval, while representations associated with inactive questions are less accessible. This could be the case because active questions represent issues that the current discourse is engaged with, and, therefore, more attention could be allocated to information associated with active questions. On the other hand, information that is not immediately relevant for resolving the active questions is less activated and less accessible. This is on par with the proposal in Grosz and Sidner (1986), where the authors suggested that "information about the objects, properties, relations, and discourse intentions that are most salient" in discourse structure becomes part of the attentional state, but anything that is irrelevant to the discourse development will be popped off the focus space. This idea is also broadly in line with previous work showing that active questions or QUDs can introduce focus to specific aspects of a target sentence and facilitate deeper processing (Cutler and Fodor, 1979; Clifton and Frazier, 2018; Sanford, 2002; Sanford & Sturt, 2002). For instance, Clifton and Frazier (2018) showed that sentences related to the QUD are processed faster. Similarly, Cutler and Fodor (1979) found that introducing a question before the target material helps comprehenders to focus on the relevant information.

In Fig. 1, in Section 3.1, we schematically illustrated how the proposal of an "active question" applies to RRCs and ARCs, especially for the type of constructions we tested in Experiment 1. In Fig. 8, we further illustrate how the difference between the ARC conditions in Experiments 1 and 2 is captured. The difference between the ARC conditions in Experiments 1 and 3 can be captured in very similar ways. Fig. 8a illustrates the ARC conditions in Experiment 1. Two discourse questions are created and processed incrementally. One is associated with the matrix clause ("What happened?") and the other is associated with the ARC ("What about the waitress?"). The ARC-associated question is removed from the stack after the ARC ends. Memory retrieval for the agreement controller starts at the matrix verb (*was/were*), and by then, the ARC-related discourse question is inactive. This leads to reduced accessibility to representations associated with the ARC, including the distractor NP (*the girl(s)*), hence reducing the likelihood of an agreement attraction effect. This is in contrast with Experiment 2, as illustrated in Fig. 8b. Again, two discourse questions are incrementally constructed, with one associated with the matrix clause and the other associated with the ARC. Since the memory retrieval of the agreement controller is initiated at a moment before the ARC is completed (i.e., at the verb *praise(s)*), the target NP (*the reviewer*) is associated with an active question ("What about the musicians?"). Also importantly, at the moment of memory retrieval, the distractor NP (*the musician(s)*) is also situated within an active discourse question (i.e., the question associated with the matrix clause "What happened?"). Both the target agreement controller (*the reviewer*) and the distractor (*the musician(s)*) are accessible for memory retrieval, resulting in a standard agreement attraction effect.



(a) ARC in Experiment 1



(b) ARC in Experiment 2

Fig. 8. Incremental construction (from left to right) of discourse questions. The box with dotted lines indicates the discourse question is no longer active. The linguistic inputs pointed with solid arrows indicate information accessible for retrieval; those pointed with dotted arrows indicate information inaccessible for retrieval. Bold-faced word = target word; italicized word = distractor; underlined word = retrieval site.

Table 15 we saw earlier also provides a summary of whether the target and distractor NPs are associated with an [+ACTIVE] or [−ACTIVE] question (see the column "Active state of discourse questions"). The ARC condition in Experiment 1 showed no agreement attraction effect, consistent with the fact that it is exactly in these conditions that the distractor NPs are associated with an [−ACTIVE] discourse question.

7.2. Discourse questions and the event model

Under our proposal, the [+ACTIVE] and [−ACTIVE] discourse questions essentially partition the discourse into distinct "domains," which results in different levels of accessibility for memory retrieval. This idea overlaps in some interesting ways with the idea of an "event" entity in Event Segmentation Theory (e.g., Radvansky & Zacks, 2017; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Under this framework, our cognitive system segments activities into events and represents the world around us using various event models. The event models are connected through causal or other types of logical relations. The specific ways these events and event models are constructed have consequences for how we perceive and remember our experience. For example, information from the current event model is more accessible than

information from previous events that are not the focus of the current attention. Therefore, it could be more difficult to retrieve information from an event after having passed the relevant event boundary.

Event segmentation has been shown to affect narrative comprehension (Bailey, Kurby, Sargent, & Zacks, 2017; Smith, Kurby, & Bailey, 2023; Zacks, Speer, & Reynolds, 2009). But it is an open question of what guides people's segmentation of events. Both linguistic and nonlinguistic physical cues have been suggested to contribute to segmenting events (Papenmeier, Maurer, & Huff, 2019, i.a.). For example, in understanding a movie scene, physical cues such as shot durations or shot transitions can be used for the purpose of segmenting event boundaries. In the current study, ARC clauses were signaled by commas. Since commas provide salient physical cues to comprehenders, it raises the question of whether the effects we observed on ARCs are discourse effects per se, or simply reflect how commas as salient physical signals can guide the segmentation of and transition between different event units.

It is true that English ARC clauses frequently co-occur with commas. For this reason, future work should also examine languages in which there is not such a correlation. But even for English, we do not think our findings can be attributed to the presence of commas alone. In particular, we note that all three experiments in the current study used commas to signal ARC clauses, and yet only in one experiment (Experiment 1) did we find that ARCs created a distinct *accessibility domain* compared to RRCs. In Experiment 2–3, ARCs and RRCs patterned in similar ways despite the fact that ARCS were still marked by commas. We suggest that discourse questions are potentially useful constructs to help delineate event boundaries as well in language comprehension. An [+ACTIVE] discourse question is an issue that the current discourse is engaged with, establishing salient boundaries from those [−ACTIVE] discourse questions. Information within the currently activated event model, encompassed by the [+ACTIVE] discourse question, is highly available for memory retrieval, and can potentially interfere with each other. But an [−ACTIVE] discourse question, similar to an inactive (or less activated) event, would be less accessible for memory retrieval.

7.3. Alternative approaches

In addition to our question-based account, several alternative accounts could potentially explain the findings of the current study. We discuss these approaches in this section.

7.3.1. Rapid loss of syntactic structure representations

Dillon et al. (2017) suggested that the syntactic form of the appositive structures, in contrast to regular relative clauses, can be lost rapidly, making it less accessible for memory retrieval. The idea of syntactic loss of appositive structure has some appeal if we assume that the appositive content will eventually become part of the given information in a comprehender's mental model (e.g., AnderBois et al., 2015), with the conceptual representation of the appositive content held in memory, while the precise structural form decays quickly.

The drawback of this account, however, is that the empirical evidence is mixed regarding whether surface structures indeed undergo a rapid forgetting process. Some work has shown evidence that syntactic forms decay fairly quickly (Lombardi & Potter, 1992; Potter

& Lombardi, 1990). For instance, Lombardi and Potter (1992) examined the robustness of maintaining the surface form of a sentence. In a sentence-recall paradigm, participants were presented with a target sentence and were instructed to recall the sentence. In between the exposure and recall phrases, participants were also presented with a distractor verb that was synonymous with the verb in the target sentence. It was found that even the brief exposure to a simple distractor verb was sufficient to disrupt recall accuracy, and this provided some evidence that the surface form of a sentence is short-lived. On the other hand, there is also work showing that surface forms may last for a long time (Gurevich, Johnson, & Goldberg, 2010; Kaschak, Kutta, & Schatschneider, 2011). For instance, Kaschak et al. (2011) found a strong structural priming effect with a long lag (even as long as a week) between the priming exposure and the target production task. Participants in their study showed a strong tendency to repeat the structure that they were exposed to even after a week from the exposure phase. It is difficult to directly compare these studies since they have used different paradigms. But it is fair to conclude that the specific constraints on when and how structural forgetting could take place are yet to be established.

7.3.2. The effect of prosody

Another promising approach for understanding the unique properties of the ARCs highlights the role of prosody. ARCs in English, as opposed to RRCs, are commonly marked with an intonation boundary before the RC, usually signaled with a pause (Dehé, 2014; Watson & Gibson, 2004). Earlier work has shown that prosodic boundaries in general can impact parsing decisions, such as resolving ambiguities, and guiding syntactic attachment preferences (Breen, 2014; Fodor, 1998, 2002; Frazier, Carlson, & Clifton, 2006; Jun & Bishop, 2015; Watson & Gibson, 2004; Wagner & Watson, 2010) and mitigating or intensifying processing difficulty (Bader, 1998). Although our testing stimuli were presented visually, we cannot exclude the effects of implicit prosody (Fodor, 1998, 2002). Previous work has shown that grouping sentences into prosodic chunks facilitates the processing of the upcoming linguistic material (Hirotani, Frazier, & Rayner, 2006; Staub, 2007), indicative of a close relationship between prosodic grouping and syntactic parsing. Particularly relevant for the current purpose, Hirotani et al. (2006) found that ARCs are read faster than the RRC counterparts, which they interpreted as a facilitation effect of integrating the upcoming linguistic material with the help of intonational grouping. Along similar lines, Kroll and Wagers (2019) suggested that prosodic grouping and syntactic parsing can go in tandem. Syntactic constituents are less likely to compete for short-term memory resources when they are in separate prosodic groupings. Appositives, therefore, are less likely to compete with the hosting matrix clause, given a clear prosodic boundary from the matrix clause, in contrast to RRCs.

We believe the connection between prosodic grouping and syntactic parsing is valid, and it is likely at work while processing ARCs. However, this approach alone is not sufficient to account for all the results in the current study. In particular, we note that in all of our experiments, there was a prosodic difference between ARCs and RRCs. The prosodic grouping account will not be able to explain why the prosodic differences between ARCs and RRCs only led to different agreement attraction effects in Experiment 1, but not in Experiments 2–3. The discourse question-based account was developed to capture the differences between the three experiments. It is entirely possible that prosodic grouping and discourse-question

tracking are two independent processes that are simultaneously at work. Whether there could be interactions between these two processes is an avenue for future work.

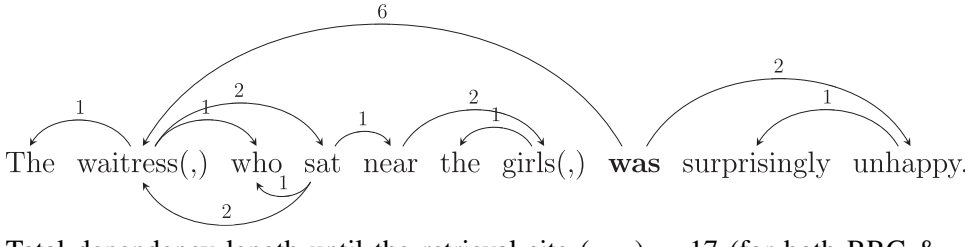
7.3.3. Dependency length difference between experiments

The structures we tested in Experiments 1–3 are not entirely the same. This leads to the potential concern that different parsing mechanisms may have been evoked in these experiments, which can independently explain the observed effects. One potential structural difference across different experiments is dependency length. Experimental work and large-scale corpus studies have shown that languages exhibit preferences for minimizing dependency length (Futrell, Mahowald, & Gibson, 2015; Gildea & Temperley, 2007, 2010; Gibson et al., 2019; Liu, 2008; Liu, Xu, & Liang, 2017; Temperley & Gildea, 2018; Yadav, Mittal, & Husain, 2022). Dependency length minimization has also been correlated with cognitive constraints such as limited working memory capacity in a way that sentences with a longer dependency incur greater processing cost than those with shorter dependency (e.g., Gibson, 1998; Grodner & Gibson, 2005). It is possible that structures with a larger count of dependency length may be more susceptible to interference effect, due to higher demand of working memory resources.

We calculated the total dependency length of the structures tested in Experiments 1 and 2. Following the Universal Dependencies (UD) project guidelines (version 2) (Nivre et al., 2020) and adopting the method from Futrell et al. (2015), we measured the total dependency length from the beginning of the target sentence up till the critical retrieval site, where the number agreement dependency is completed (see (9)). The dependency length was calculated by summing all the closed dependencies on the left-hand side of the retrieval site. Most importantly, the RRC and ARC structures we tested do not differ in their total count of dependency length.

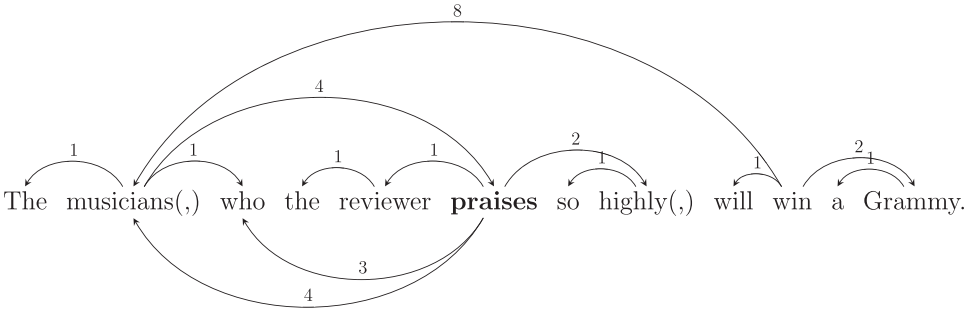
(9) Dependency length calculation

a. Experiment 1



Total dependency length until the retrieval site (*was*) = 17 (for both RRC & ARC)

b. Experiment 2



Total dependency length until the retrieval site (*praises*) = 15 (for both RRC & ARC)

The summed dependency length until the retrieval site is 17 in Experiment 1, and 15 in Experiment 2. The difference between the two experiments is small. If a slightly higher count of total dependency length could nonetheless be more demanding on working memory and potentially result in a higher degree of memory interference, we might have expected more agreement attraction effect in Experiment 1 than in Experiment 2. This prediction was not borne out. The two experiments exhibited a similar number agreement attraction effect in the RRC conditions; and for the ARCs, it was Experiment 2 instead of Experiment 1 that exhibited an attraction effect.

7.3.4. Different position of the distractor

Another structural difference between Experiment 1 and Experiments 2–3 concerns the position of the distractor. In Experiment 1, the distractor intervenes between the subject-verb agreement dependency, but in Experiments 2–3, the distractor does not intervene. Staub (2010) argued that depending on the position of the distractor, there may be different types of number agreement attraction effects. In particular, in Staub (2010), the agreement attraction effect in a distractor-non-intervening construction was mostly driven by trials with long reading times, that is, the effect was primarily present in the right tail of the RT distribution. On the other hand, in a distractor-intervening construction, the agreement attraction effect was more evenly distributed in the entire RT distribution. This difference was considered to be indicative of a qualitatively distinct processing mechanism underlying the agreement attraction effect in different constructions.

If intervening and nonintervening distractors are associated with distinct parsing mechanisms, this could potentially challenge our discourse question account of the current findings. To investigate this, we conducted an ex-Gaussian analysis on the observed agreement attraction effects, following the methods in Staub (2010) and Huang and Dillon (2023). Ex-Gaussian modeling allows us to estimate whether the agreement attraction effects we observed in Experiments 1–3 reside in the shift of the mean between the RT distribution of the target condition and its control condition, or by a shift in skewness (i.e., an effect in the tail of the RT distribution), or a combination of both. More importantly, since the RRC conditions (in particular from Experiments 1 and 2) in the current study were similar to the stimuli used in Staub (2010), we will use the results from the RRC conditions to

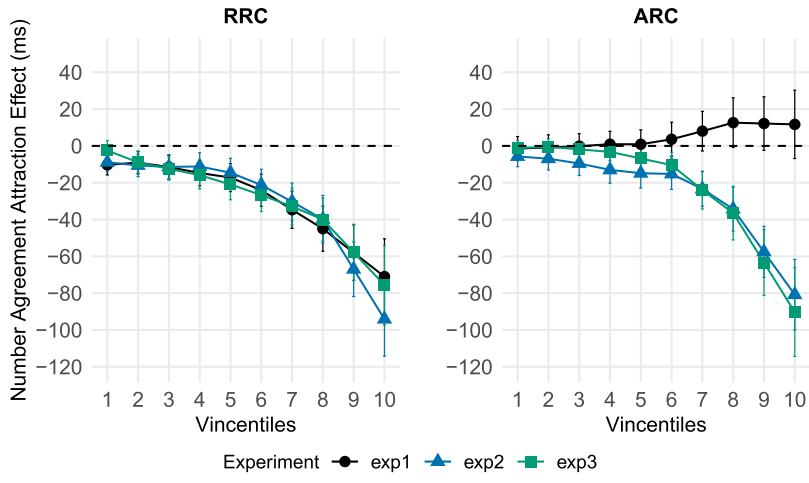


Fig. 9. Vincentile plot of number agreement attraction effect. The effect was calculated by subtracting the ungrammatical singular condition from the ungrammatical plural condition. Error bars represent standard errors.

evaluate whether the agreement attraction effects in the current study show qualitatively distinct patterns based on the intervening and nonintervening distractor contrast discussed in Staub (2010). Furthermore, for a comparison between RRC and ARC conditions, we also fitted ex-Gaussian models to the ARC data as well. For the sake of space, we present the details of the ex-Gaussian modeling and its outcome in Appendix B; and we only visually illustrate the findings below through vincentile plots. But to preview the results presented in Appendix B, contrary to Staub (2010), we did not find clear evidence that the position of the distractor triggered distinct processing mechanisms.

The vincentile plots in Fig. 9 were generated following the procedures outlined in Staub (2010). Focusing on the conditions most relevant for the agreement attraction effect, Fig. 9 only targeted the ungrammatical conditions at the spillover region. For each clause type (RRC or ARC), we divided the data points for each subject into vincentiles, with the fastest 10% of reading times as vincentile 1, the next fastest group as vincentile 2, and so on. Subsequently, we computed the agreement attraction effect by subtracting the singular ungrammatical condition from the plural ungrammatical condition for each subject and each vincentile. Then, an averaged agreement attraction effect was calculated for each vincentile. The Y-axis values smaller than 0 ms in Fig. 9 indicate a standard number agreement attraction effect.

As shown by the results for the RRC conditions (left panel in Fig. 9), the size of the agreement attraction effect was much larger in the slower RT range, suggesting that the skewness difference instead of the mean difference between the plural versus the singular distractor condition is more responsible for the observed attraction effect. Most crucially, the pattern appears consistent across all three experiments, suggesting that the different positions of the distractor did not necessarily lead to distinct processing mechanisms in the current study. For the ARC conditions (right panel in Fig. 9), consistent with the results reported in the result section, Experiment 1 did not show a robust agreement attraction effect. Furthermore,

Experiments 2–3, both containing nonintervening distractors, showed similar patterns. Much like the agreement attraction effects in the RRCs, the attraction effect in the ARCs was most prominent in the slower RT range.

7.4. Limitations and future directions

Before we conclude, there are a few limitations of the current study that are worth noting. First, the current study only examined a very specific type of appositives—the nonrestrictive (appositive) relative clauses. We used commas to cue people the distinction between RRCs and ARCs. Previous research has demonstrated that readers do attend to commas despite their seemingly trivial visual salience (Angele, Gutiérrez-Cordero, Perea, & Marcet, 2024; Hirotani et al., 2006; Stine-Morrow et al., 2010; Warren, White, & Reichle, 2009). But it is possible that there is individual variability in their perception of the comma cue, and there may also be grammatical variations regarding the effectiveness of using commas to diagnose ARCs. The variations may be partly driven by factors like age and reading habits (e.g., Stine-Morrow et al., 2010), which we did not investigate in the current study. Future work on other types of appositive structures would be necessary in order to establish broader empirical consensus. We note that there have been some promising results from a number of studies, such as Ng and Husband (2017) and McInnerney and Atkinson (2020), that revealed findings similar to the current Experiment 1, despite using a different type of appositives (*the one who/that*…).

Another limitation of the current study is that in Experiment 1, the regular mixed effects model did not detect a significant three-way interaction between clause, distractor, and grammaticality. Additionally, in Section 6.4, when the critical ARC conditions of all three experiments were compared together, we also did not obtain a significant three-way interaction between distractor, grammaticality, and experiments. It is possible that the current sample size is not large enough to achieve sufficient statistical power for the complex interaction effect. In particular, in both analyses above, the relevant significant effects were obtained when we focused only on the ungrammatical trials. It has been shown in previous work that although the number feature interference effect is robust in ungrammatical sentences, it is more variable in grammatical sentences (e.g., Jäger et al., 2017; Laurinavichyute & von der Malsburg, 2024). This may have led to greater noise in the data, making it more difficult to detect a three-way interaction involving the grammaticality predictor. Another interesting possibility is that the standard analysis procedures we adopted for analyzing self-paced reading time may need to be revisited. The ex-Gaussian modeling results we presented in Section 7.3.4 and Appendix B are informative in this regard. First of all, these additional analyses were highly consistent with our original analyses in revealing a difference between RRCs and ARCs in Experiment 1, while finding similar patterns between the two clause types in Experiments 2 and 3. But more importantly, we showed that across all three experiments, when agreement attraction arises, the effect predominately resided in the shift of the skewness of the RT distribution, instead of a shift in the mean of the RT distribution. In other words, the attraction effect was driven largely by the subset of trials that had longer reading times. The standard analysis procedures for reading time measures, however, have focused on modeling the differences in means. This raises the question of whether the distributional properties of the agreement attraction effect, or maybe self-paced reading studies in general, can be adequately

captured by traditional data analysis tools, especially when a complex interaction effect needs to be detected. A more detailed exploration of this problem is beyond the scope of the current study. For future work, it would be valuable to conduct preregistered replication studies with a larger sample size (e.g., Wicherts et al., 2016). It is also worth exploring new analysis techniques that can better capture the natural distribution properties of agreement attraction.

Finally, our question-based approach to discourse structure is applied to the empirical case of the agreement attraction effect. More work is needed to establish whether this approach has general appeal across different linguistic dependencies. A small number of studies have compared RRCs versus ARCs containing other types of dependencies. Ng and Husband (2017), for instance, examined the negative polarity item (NPI) licensing in appositives and RRCs. Kim and Xiang (2023) compared pronoun resolution that required locating an antecedent from either ARCs or RRCs. The findings were mixed. Ng and Husband (2017) did not observe a contrast between RRCs and ARCs when NPI licensing was concerned, but Kim and Xiang (2023) showed that pronoun resolution was sensitive to the clause type (discourse structure) distinction. It is premature to draw any strong conclusions based on the limited set of findings. We need future work that systematically examines how the effect of incremental discourse update interacts with various different types of structures.

8. Conclusion

Results from three self-paced reading experiments suggest that the well-known agreement attraction effect is constrained by discourse structure. More importantly, the impact of discourse structure was not realized simply by representing the relevant discourse information as static features. Instead, to capture all the results, we need to pay close attention to the incremental update of the overall discourse structure, specifically the active status of discourse questions at any given moment of a discourse. The current findings have implications for the memory retrieval mechanism, suggesting that memory retrieval is constrained by the active status of discourse questions.

Acknowledgments

We are grateful to the audience at the 35th Annual Conference on Human Sentence Processing, and the lab members at the Language Processing Lab at the University of Chicago for discussion and comments. We appreciate the anonymous reviewers and the editor for suggestions and feedback. Special thanks to Kuan-Jung Huang for his help with data analysis. All shortcomings are our own.

Funding information

This project was supported by an NSF DDRI Grant (BCS-2214437).

Conflict of interest statement

The authors have no conflicts of interest to declare.

Data availability statement

Stimuli, data, and scripts for data analysis and visualization have been made publicly available at the Open Science Framework (OSF): https://osf.io/rsdp5/

Notes

1 It is not always the case that ARCs are in a subordinating relation relative to the rest of the discourse (Göbel, 2019; Göbel, Frazier, & Clifton, 2021; Hunter & Asher, 2016; Jasinskaja, 2016; Syrett & Koev, 2015). But in the current work, we primarily focus on the constructions like (1), which exhibits a more robust contrast between ARC and RRC structures in terms of their distinct discourse status.

2 The specific syntactic analysis for ARCs is subject to debate (de Vries, 2006; Griffiths & de Vries, 2013; Jackendoff, 1977). See de Vries (2006) and McInnerney (2022) for an overview.

3 In contrast to the linguistic concept of Question under Discussion (QUD) (Roberts, 2012), the term "discourse question" is used more broadly here, encompassing questions related to both main utterances and side-commentary content.

4 In the current study, we only focus on the facilitatory interference/attraction effect in ungrammatical sentences. We did not observe any effect in grammatical sentences. But it remains an open question whether agreement attraction is asymmetric in grammatical versus ungrammatical sentences (Dillon, Mishler, Sloggett, & Phillips, 2013; Hammerly, Staub, & Dillon, 2019; Jäger, Engelmann, & Vasishth, 2017; Lago, Shalom, Sigman, Lau, & Phillips, 2015; Wagers et al., 2009).

5 Also see Vasishth (2021) for the benefits of using nested contrasts; and see Patil, Vasishth, and Lewis (2016) for a similar example using nested contrast for studying the interference effect.

6 Initially, a total of 96 participants were recruited. At a reviewer's request, 24 participants were additionally recruited to match the number of participants across experiments. The additional participants did not change the results of Experiment 2.

7 Initially, a total of 96 participants were recruited. At a reviewer's request, 24 participants were additionally recruited to match the number of participants across experiments. The additional participants did not change the results of Experiment 3.

8 The ex-Gaussian distribution had three parameters, μ (mean), τ (skewness), and σ (standard deviation). We assumed there was no distractor effect on σ.

References

AnderBois, S., Brasoveanu, A., & Henderson, R. (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics*, *32*(1), 93–138.

Angele, B., Gutiérrez-Cordero, I., Perea, M., & Marcet, A. (2024). Reading(,) with and without commas. *Quarterly Journal of Experimental Psychology*, *77*(6), 1190–1200.

Asher, N. (1993). *Reference to abstract objects in English: A philosophical semantics for natural language metaphysics (Studies in Linguistics and Philosophy)*. Dordrecht: Kluwer.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge: Cambridge University Press.

Bach, K. (1999). The myth of conventional implicature. *Linguistics and Philosophy*, *22*, 327–366.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht: Kluwer

Bailey, H. R., Kurby, C. A., Sargent, J. Q., & Zacks, J. M. (2017). Attentional focus affects how events are segmented and updated in narrative reading. *Memory & Cognition*, *45*, 940–955.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Birner, B. J., & Ward, G. L. (1998). *Information status and noncanonical word order in English* (volume 40). Philadelphia, PA: John Benjamins.

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, *8*(2), 37–50.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bruce, K. B., & Farkas, D. F. (2007). Context structure for dialogues. *Unpublished manuscript, University of California, Santa Cruz*.

Büring, D. (2003). On d-trees, beans, and b-accents. *Linguistics and Philosophy*, *26*, 511–545.

Canac-Marquis, R., & Tremblay, M. (1998). The wh-feature and the syntax of restrictive and nonrestrictive relatives in French and English. In J. Lema & E. Treviño (Eds.), *Theoretical analyses on Romance languages* (pp. 127–141). Amsterdam: John Benjamins

Clifton, Jr, C., & Frazier, L. (2018). Context effects in discourse: The question under discussion. *Discourse Processes*, *55*(2), 105–112.

Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language*, *102*, 16–27.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59.

de Vries, M. (2006). The syntax of appositive relativization: On specifying coordination, false free relatives, and promotion. *Linguistic Inquiry*, *37*(2), 229–270.

Dehé, N. (2014). *Parentheticals in spoken English: The syntax-prosody relation*. Cambridge: Cambridge University Press.

Dehé, N., & Kavalova, Y. (2007). Parentheticals: An introduction. In N. Dehè & Y. Kavalova (Eds.), *Parentheticals*, volume 106 of *Linguistics Today* (pp. 1–24). Amsterdam: John Benjamins

Dever, J. (2001). Complex demonstratives. *Linguistics and Philosophy*, *24*, 271–330.

Dillon, B., Clifton, Jr, C., & Frazier, L. (2014). Pushed aside: Parentheticals, memory and processing. *Language, Cognition and Neuroscience*, *29*(4), 483–498.

Dillon, B., Clifton, Jr, C., Sloggett, S., & Frazier, L. (2017). Appositives and their aftermath: Interference depends on at-issue vs. not-at-issue status. *Journal of Memory and Language*, *96*, 93–109.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103.

Duff, J., Anand, P., Brasoveanu, A., & Rysling, A. (2023). Pragmatic representations and online comprehension: Lessons from direct discourse and causal adjuncts. *Glossa Psycholinguistics*, *2*(1), 3, 1–52. https://doi.org/10.5070/G6011198

Emonds, J. (1979). Appositive relatives have no properties. *Linguistic Inquiry*, *10*, 211–243.
Fabb, N. (1990). The difference between English restrictive and nonrestrictive relative clauses. *Journal of Linguistics*, *26*(1), 57–77.
Farkas, D. F., & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, *27*(1), 81–118.
Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*, 285–319.
Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In B. Bel, & I. Marlien (Eds.), *Proceedings of Speech Prosody* (pp. 83–90). Aix-en-Provence, France: Laboratoire Parole et Langage, SProSIG.
Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language*, *56*(3), 357–383.
Frazier, L., Carlson, K., & Clifton, Jr, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249.
Frazier, L., & Clifton, Jr, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93–126.
Frazier, L., & Clifton, Jr, C. (2005). The syntax-discourse divide: Processing ellipsis. *Syntax*, *8*(2), 121–174.
Frazier, L., Dillon, B., & Clifton, Jr, C. (2018). Together they stand: Interpreting not-at-issue content. *Language and Speech*, *61*(2), 199–226.
Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341.
Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.
Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. P. Marantz, Y. Miyashita, & O. Wayne (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.
Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407.
Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, *7*(1), 55–78.
Gildea, D., & Temperley, D. (2007). Optimizing grammars for minimum dependency length. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (pp. 184–191).
Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science*, *34*(2), 286–310.
Göbel, A. (2019). Final appositives at the right frontier: An experimental investigation of anaphoric potential. In M. T. Espinal, E. Castroviejo, M. Leonetti, L. McNally, & C. Real-Puigdollers (Eds.), *Proceedings of Sinn und Bedeutung 23* (pp. 451–467). Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès).
Göbel, A., Frazier, L., & Clifton, Jr, C. (2021). Tense and discourse structure: The timeline hypothesis. *Discourse Processes*, *58*(9), 852–868.
Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411.
Griffiths, J., & de Vries, M. (2013). The syntactic integration of appositives: Evidence from fragments and ellipsis. *Linguistic Inquiry*, *44*(2), 332–344.
Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261–290.
Grosz, B., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.
Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition*, *2*(1), 45–78.
Haegeman, L. (2008). Parenthetical adverbials: The radical orphanage approach. In Benjamin Shaer, Philippa Cook, Werner Frey & Claudia Maienborn (Eds.), *Dislocated elements in discourse* (pp. 339–355). London: Routledge.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)* (volume 2, pp. 1–8). Association for Computational Linguistics.

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, *110*, 70–104.

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*(3), 425–443.

Hobbs, J. R. (1985). *On the coherence and structure of discourse*. Stanford, CA.

Huang, K.-J., & Dillon, B. (2023). An infrequent, large cost underlies garden path effects: An RT distribution approach. In *The 36th Annual Conference on Human Sentence Processing (HSP2023)*, University of Pittsburgh, PA. March 9–11 (plenary talk).

Hunter, J., & Asher, N. (2016). Shapes of conversation and at-issue content. In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (Eds.), *Semantics and linguistic theory* (volume 26, pp. 1022–1042).

Jackendoff, R. (1977). *$\bar{X}$ syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339.

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, *111*, 104063.

Jasinskaja, K. (2016). Not at issue any more. *Unpublished manuscript, University of Cologne*.

Jun, S.-A., & Bishop, J. (2015). Priming implicit prosody: Prosodic boundaries and individual differences. *Language and Speech*, *58*(4), 459–473.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147.

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, *39*, 381–388.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kim, S. J., & Xiang, M. (2023). Memory encoding/retrieval is sensitive to discourse status: Through the lens of pronoun resolution. In *The 36th Annual Conference on Human Sentence Processing (HSP2023)*, University of Pittsburgh, PA. March 9–11 (plenary talk).

Koev, T. (2022). *Parenthetical meaning*. Oxford Studies in Semantics and Pragmatics. Oxford: Oxford University Press.

Kratzer, A., & Heim, I. (1998). *Semantics in generative grammar* (volume 1185). Oxford: Blackwell.

Kroll, M., & Wagers, M. (2019). Working memory resource allocation is not modulated by clausal discourse status. *Unpublished manuscript, University of California, Santa Cruz*.

Kwon, N., Ong, D., Chen, H., & Zhang, A. (2019). The role of animacy and structural information in relative clause attachment: Evidence from Chinese. *Frontiers in Psychology*, *10*, 1576.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language*, *82*, 133–149.

Lascarides, A., & Asher, N. (2008). Segmented discourse representation theory: Dynamic semantics with discourse structure. In Harry Bunt & Reinhard Muskens (Eds.), *Computing meaning* (volume 3, pp. 87–124). Dordrecht: Springer.

Laurinavichyute, A., & von der Malsburg, T. (2024). Agreement attraction in grammatical sentences and the role of the task. *Journal of Memory and Language*, *137*, 104525.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*, 339–359.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159–191.

Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, *21*, 171–193.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*(6), 713–733.

McCawley, J. D. (1982). Parentheticals and discontinuous constituent structure. *Linguistic Inquiry*, *13*(1), 91–106.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91.

McInnerney, A. (2022). Parenthetical niching: A third-factor phonosyntactic analysis. *Syntax*, *25*(3), 379–415.

McInnerney, A., & Atkinson, E. (2020). Syntactically unintegrated parentheticals: Evidence from agreement attraction. In *The 33rd Annual CUNY Human Sentence Processing*. Amherst, MA: University of Massachusetts Amherst. March 19–21 (oral presentation).

Murray, S. E. (2014). Varieties of update. *Semantics and Pragmatics*, *7*(2), 1–53.

Ng, A., & Husband, M. (2017). Interference effects across the at-issue/not-at-issue divide: Agreement and NPI licensing. In *The 30th Annual CUNY Human Sentence Processing*. Cambridge, MA: MIT. March 30–April 1 (poster presentation).

Nicenboim, B., Schad, D. J., & Vasishth, S. (2023). Introduction to Bayesian data analysis for cognitive science. https://vasishth.github.io/bayescogsci/book/.

Nivre, J., De Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., Schuster, S., Tyers, F., & Zeman, D. (2020). Universal Dependencies v2: An evergrowing multilingual treebank collection. *arXiv preprint arXiv:2004.10643*.

Onea, E., & Ott, D. (2022). Nominal appositives in grammar and discourse. *Language*, *98*(2), 359–391.

Ott, D. (2016). Ellipsis in appositives. *Glossa: a journal of general linguistics*, *1*(1), 34, 1–46. http://dx.doi.org/10.5334/gjgl.37

Papenmeier, F., Maurer, A. E., & Huff, M. (2019). Linguistic information in auditory dynamic events contributes to the detection of fine, not coarse event boundaries. *Advances in Cognitive Psychology*, *15*(1), 30.

Parker, D., & An, A. (2018). Not all phrases are equally attractive: Experimental evidence for selective agreement attraction effects. *Frontiers in Psychology*, *9*, 1566.

Partee, B. (1975). Montague grammar and transformational grammar. *Linguistic Inquiry*, *6*(2), 203–300.

Patil, U., Vasishth, S., & Lewis, R. L. (2016). Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology*, *7*, 329.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, *12*(5–6), 601–638.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633–654.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford: Oxford University Press.

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133–140.

Riester, A. (2019). Constructing QUD trees. In M. Zimmermann, K. von Heusinger, & E. Onea (Eds.), *Questions in discourse* (volume 2, pp. 163–192). Leiden: Brill.

Roberts, C. (2004). Context in dynamic interpretation. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 197–220). Oxford: Blackwell

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*(6), 1–69.

Ross, J. R. (1967). *Constraints on variables in syntax*. PhD thesis, MIT, Cambridge, MA.
Safir, K. (1986). Relative clauses in a theory of binding and levels. *Linguistic Inquiry*, *17*(4), 663–689.
Sanford, A. J. (2002). Context, attention and depth of processing during interpretation. *Mind & Language*, *17*(1–2), 188–206.
Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382–386.
Slioussar, N., & Malko, A. (2016). Gender agreement attraction in Russian: Production and comprehension evidence. *Frontiers in Psychology*, *7*, 1651.
Smith, M. E., Kurby, C. A., & Bailey, H. R. (2023). Events shape long-term memory for story information. *Discourse Processes*, *60*(2), 141–161.
Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 550–569.
Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.
Stine-Morrow, E. A., Shake, M. C., Miles, J. R., Lee, K., Gao, X., & McConkie, G. (2010). Pay now or pay later: Aging and the role of boundary salience in self-regulation of conceptual integration in sentence processing. *Psychology and Aging*, *25*(1), 168.
Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*(5), 882–888.
Syrett, K., & Koev, T. (2015). Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics*, *32*(3), 525–577.
Temperley, D., & Gildea, D. (2018). Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics*, *4*, 67–80.
Truckenbrodt, H. (2015). Intonation phrases and speech acts. In M. Kluck, D. Ott, & M. de Vries (Eds.), *Parenthesis and ellipsis: Cross-linguistic and theoretical perspectives*, Studies in Generative Grammar (pp. 301–349). Berlin: Mouton de Gruyter.
Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316.
Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263.
Vasishth, S. (2021). A common mistake in psychology and psycholinguistics: Part 2. https://vasishth-statistics.blogspot.com/2021/08/a-common-mistake-in-psychology-and_13.html.
Voeten, C. C. (2022). *buildmer: Stepwise elimination and term reordering for mixed-effects regression*. R package version 2.6.
Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.
Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304.
Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7–9), 905–945.
Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112.
Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E–Z Reader. *Cognition*, *111*(1), 132–137.
Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–755.
Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R. C., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid *p*-hacking. *Frontiers in Psychology*, *7*, 1832.
Yadav, H., Mittal, S., & Husain, S. (2022). A reappraisal of dependency length minimization as a linguistic universal. *Open Mind*, *6*, 147–168.

Yadav, H., Paape, D., Smith, G., Dillon, B. W., & Vasishth, S. (2022). Individual differences in cue weighting in sentence comprehension: An evaluation using Approximate Bayesian Computation. *Open Mind*, *6*, 1–24.
Yadav, H., Smith, G., Reich, S., & Vasishth, S. (2023). Number feature distortion modulates cue-based retrieval in reading. *Journal of Memory and Language*, *129*, 104400.
Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307.
Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273.

Appendix A

Figs. A1–A3 present the individual data points by subject concerning the interference

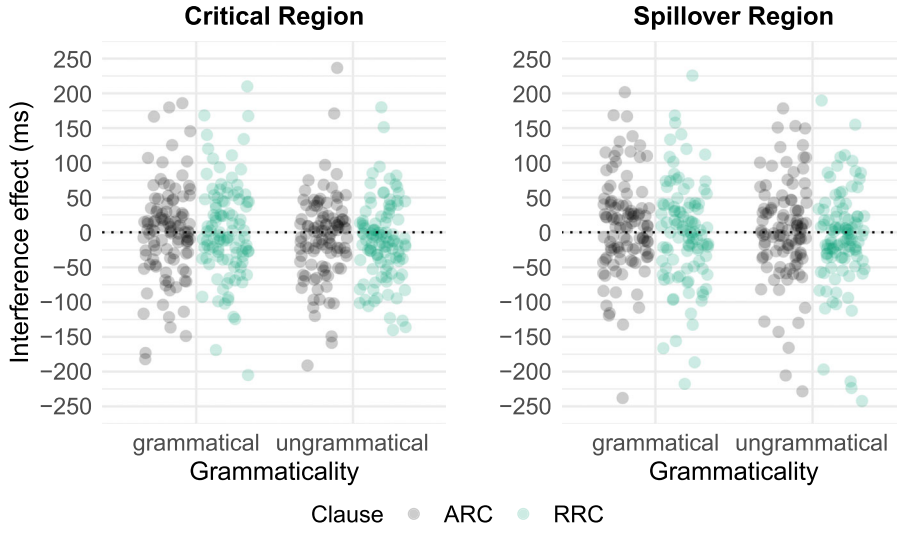


Fig. A1. Interference effect by subject in Experiment 1.

effect. The interference effect was calculated by subtracting the untransformed mean reading time of the singular distractor condition from the plural distractor condition. Data points with a negative value in the ungrammatical condition show the standard number agreement attraction effect.

Appendix B

To examine whether the agreement attraction effects we observed in Experiments 1–3 were driven by the mean difference between the target condition and its control condition, or by a shift in skewness (i.e., an effect from the tail of the distribution), or a combination of both, we fitted ex-Gaussian models to the data. Ex-Gaussian modeling allows us to separately estimate effects from the shift of the mean RT (the μ parameter) and effects from the shift of

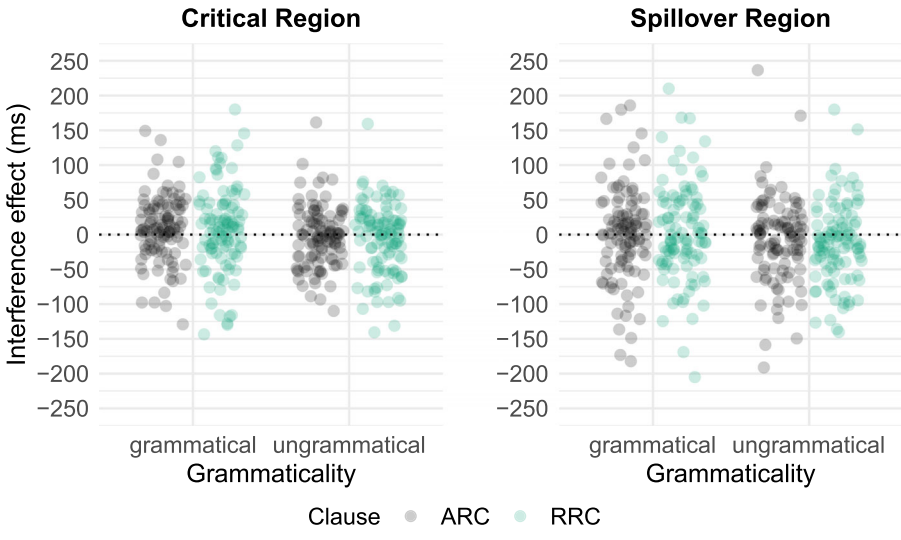


Fig. A2. Interference effect by subject in Experiment 2.

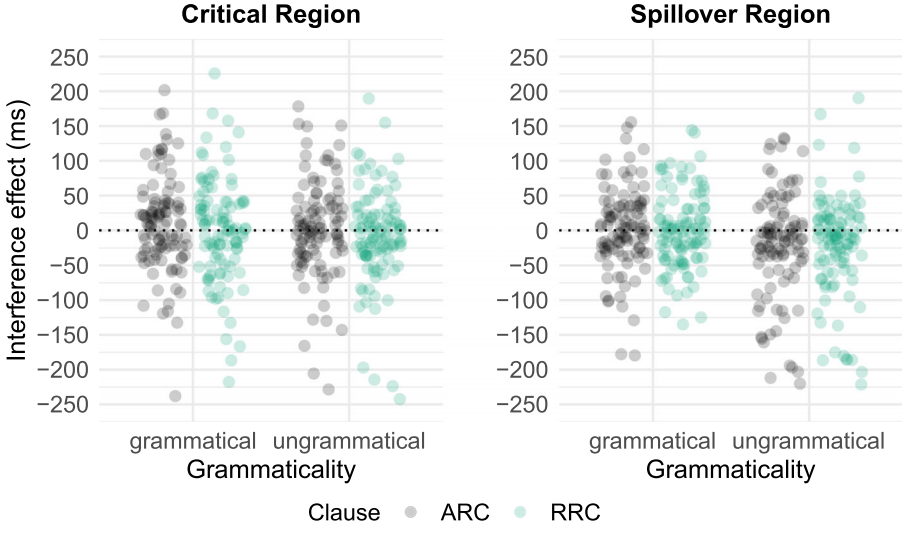


Fig. A3. Interference effect by subject in Experiment 3.

skewness (the τ parameter). Linear mixed-effects models with an ex-Gaussian distribution were implemented in R (R Core Team, 2022) using the `brms` library (Bürkner, 2017). As our primary interest was the number agreement attraction effect, which only appeared in the ungrammatical conditions, we analyzed the data from the ungrammatical conditions at the spillover region only. A separate model was fit for each clause type and experiment, where the raw reading time (ms) was the dependent variable, and the distractor was included as the

Table B1
The average agreement attraction effect, and the mean (μ) and skewness (τ) effects in the ungrammatical condition at the spillover region

	Exp. 1		Exp. 2		Exp. 3	
	RRC	ARC	RRC	ARC	RRC	ARC
Average attraction effect	−25.71	−2.17	−22.08	−23.60	−11.04	−11.55
τ effect on attraction	−27.31	3.41	−27.53	−31.52	−25.68	−20.24
μ effect on attraction	1.6	−5.58	5.45	7.92	14.64	8.69

Note. The attraction effect is calculated by the difference between the two distractor conditions(= plural−singular). Estimates are calculated in raw reading times (ms).

fixed effect (treatment coded singular = 0; plural = 1). Random effects of item and subject were also included in the model.[8]

We converted the parameter estimates in the model outcome into milliseconds, and a summary of the findings is presented in Table B1. The averaged attraction effect presented in the table is the overall averaged reading time difference between the two distractor conditions (= plural−singular distractor). This effect is further decomposed into an effect of τ (distribution shift in skewness) and an effect of μ (distribution shift in mean); the averaged effect is the sum of the effects in these two parameters.

For Experiment 1, the results in Table B1 confirmed our conclusions in Section 4.2 that the attraction effect is present in the RRC but not in the ARC conditions (attraction effect in RRC: −25.71 ms; ARC: −2.17 ms). More importantly, we see that the attraction effect in the RRC conditions is entirely driven by the τ (around −27.31 ms). For Experiment 2, the magnitude of the attraction effect is similar between RRCs and ARCs (on average −22 and −24 ms); and more importantly, for both RRCs and ARCs, the attraction effect is driven by the τ effect (RRC: −28 ms, ARC: −32 ms). The estimates for the μ effect are small and positive for both clause types (RRC: 5 ms, ARC: 8 ms), essentially showing no facilitatory agreement attraction effect in the μ. It is worth noting that the experimental constructions in Experiments 1 and 2 are similar to those tested in Staub (2010). But contrary to Staub (2010), we found no clear evidence that the position of the distractor (i.e., whether the distractor intervenes in between the subject-verb dependency) triggers a different agreement attraction effect.

Experiment 3 also contained nonintervening distractors, similar to Experiment 2. Interestingly, although the effects were by and large consistent between these two experiments, there were some nuanced differences as well. In particular, similar to Experiment 2, the attraction effects in Experiment 3 were primarily driven by the τ, but the τ effects were slightly smaller than Experiment 2. The effects in the μ were again the opposite of a standard attraction effect, but the μ effect was greater in Experiment 3 than Experiment 2, especially with the RRC conditions (14.64 vs. 5.45 ms). It is possible that the slight difference in Experiment 3 could be a consequence of the different structures used. Specifically, the relative clauses in Experiment 3 were at the sentence-final position, whereas in Experiment 2 (and Experiment 1 too), the relative clauses were sentence-medial. We leave this question for future work.