Detecting the Effects of Antecedent Complexity on VPE Resolution

Jiayuan Yue and Ming Xiang

1 Introduction

Sentences containing Verb-Phrase Ellipsis (VPE), as exemplified in (1), consist of a VP antecedent and a subsequent omission of the VP. In (1), "Mary did too" means that Mary read a book too. As a reader reads the sentence, they first form a mental representation of the VP antecedent "read a book". Then upon the site of ellipsis, they need to access this mental representation and retrieve the information to interpret the VPE. Therefore, studying the process of VPE resolution is highly informative for us to understand the memory retrieval mechanism in human sentence processing. This paper explores the following research question: Could the representational complexity of the VP antecedent affect the difficulty of its subsequent retrieval at the ellipsis site?

(1) John read a book and Mary did too.

A large body of research has studied similar questions in filler-gap dependencies other than ellipsis, most notably in relative clauses (Hofmeister 2011, Hofmeister and Vasishth 2014, Troyer et al. 2016, Karimi et al. 2020, but see Lam and Xiang 2022). In self-paced reading experiments, Hofmeister (2011) found that in relative clauses like (2), a complex head noun phrase "an alleged Venezuelan communist" leads to a longer reading time at "communist" in the encoding phase (i.e. when the head noun is first encountered), but a faster reading at the region after the verb "banned", when the head noun needs to be retrieved from working memory. This observation suggests that the representational complexity of the more elaborate target NP reduces the cost of its retrieval in later sentence processing. Subsequent experiments attribute this retrieval facilitation effect to the distinctiveness of a semantically elaborate target and thus the reduction of similarity-based interference from other NP representations in memory (Hofmeister and Vasishth 2014).

(2) a. Simple: It was *a communist* who the members of the club banned from ever entering the premises.
    b. Complex: It was *an alleged Venezuelan communist* who the members of the club banned from ever entering the premises.

Some other work has attributed the complexity effect to the time spent maintaining or expecting a representation (Karimi et al. 2020). The time period of maintaining the representation would be accompanied by heightened attention on reading the target NP, which facilitates later retrieval. Karimi et al. (2020) created complex items by adding masking Korean characters in the middle of an NP like "the bear". "The" is a cue that predicts an upcoming head noun. The Korean characters did not increase the semantic distinctiveness of the NP to non-Korean speakers, but they introduced additional time spent maintaining the NP representation. The results showed a speedup at a site where the target NP needed to be retrieved. Such findings were better explained by the time/attention account than the distinctiveness account.

Besides relative clauses, pronominal resolutions also show the complexity effect (Karimi et al. 2018). A sentence like "the actor walked away from the cameraman" provides two possible referents for a subsequent pronoun, as in "after a while, he realized it was getting late and took a taxi home." Karimi et al. (2020) tested whether the complexity of the antecedents in the first sentence affects the pronoun resolution in the second sentence. The antecedent complexity was manipulated by adding modifiers to the relevant NPs, such as "the actor who was visibly upset walked away from the cameraman who was critical of the show." The results from an EEG experiment showed that the complex referents, relative to their simpler counterparts, weakened the Nref effect – an ERP signal of referential processing difficulty. Therefore, semantic complexity also facilitates memory retrieval in pronominal resolution.

Contrary to the broad literature on various filler-gap dependencies, previous studies on VPE found no significant effect of the complexity of the antecedent on the processing time at the site of ellipsis (Frazier and Clifton 2000, 2001, Martin and McElree 2008, Paape et al. 2017). The only exception is an early finding by Murphy (1985), who tested sentences like those in (3). (3a) was the simple condition with a short VP antecedent, while (3b) was the complex condition with a longer antecedent. Participants performed a reading task where they read one sentence at a time and click on to the next sentence once they think they fully understand the previous one. The reading time was found to be longer in the complex condition than in the simple condition. Therefore, Murphy proposed a "copy" mechanism of VPE resolution, where a reader copies the contents of the antecedent into the ellipsis site. A complex antecedent would therefore take longer to copy.

(3) a. Simple: Jimmy swept the floor. Later, his uncle did too.
b. Complex: Jimmy swept the tile floor behind the chairs free of hair and cigarettes. Later, his uncle did too.

However, the materials in Murphy's (1985) experiment contain an attachment ambiguity (Tanenhaus and Carlson 1990). For instance, it is unclear what phrase "free of hair and cigarettes" attaches to. The increase in reading time in the complex condition could be due to the difficulty of resolving the attachment ambiguity, rather than the difficulty of retrieving a longer antecedent. Frazier and Clifton (2000, 2001) conducted a self-paced reading experiment and found no effect of antecedent complexity on VPE resolution. The complexity manipulation of the VP antecedent in (4a-b) did not affect the reading time of the elliptical sentence (4c), contrary to Murphy's proposal of a copy mechanism. Instead, Frazier and Clifton (2001) proposed the copy α mechanism. The ellipsis only generates a shallow copy of the antecedent, and the copying in the interpretation is cost-free.

(4) a. Simple: Sarah left her boyfriend last May.
b. Complex: Sarah got up the courage to leave her boyfriend last May.
c. Ellipsis: Tina did too.

In self-paced reading experiments of elliptical sentences, a common concern is that participants may not thoroughly retrieve and integrate the VP antecedent, making the complexity manipulation ineffective. The reader might have just been pressing the space button mechanically and does not fully understand the meaning of the ellipsis before going beyond the relevant chunks. If this is the case, then the reading time at the ellipsis site would not be indicative of the processing difficulty of the VPE, because the VPE is not adequately processed in the first place. This is an important concern for reading tasks of elliptical sentences, and we must control for such shallow processing before making specific conclusions about ellipsis resolution.

Martin and McElree (2008) conducted an experiment that required participants to judge the semantic compatibility of the elliptical sentence as a comprehension check. Participants had to integrate and comprehend the VPE to be able to judge its semantic compatibility. The experiment used a speed-accuracy tradeoff task (SAT). The SAT paradigm has the advantage in teasing apart processing speed from processing accuracy, when these two aspects are normally conflated in regular behavioral measures. The example in (5) demonstrates the complexity manipulation of the VP antecedent. The complex VP in (5b) contains a more elaborate NP compared to the simpler VP in (5a). The semantic compatibility of the elliptical sentence was manipulated by changing the subject NP of the ellipsis site. An overworked student can understand Roman mythology but an overly worn book cannot understand anything. Participants were instructed to indicate their judgment as soon as possible after the last phrase of the sentence "did not" was presented.

(5) a. Simple: The history professor understood Roman mythology, but the principal was displeased to learn that the [overworked students attending summer session / overly worn book used in summer session] did not.
b. Complex: The history professor understood Rome's swift and brutal destruction of Catharge, but the principal knew the [overworked students attending summer session / overly worn books used in summer session] did not.

The results showed no effect of antecedent complexity on the processing speed of the ellipsis site. Again, this result did not support Murphy's (1985) copy mechanism. Instead of assuming a cost-free copy mechanism (e.g. the cost-free *copy α* in Frazier and Clifton 2001), Martin and McElree (2008) argued that copying, as an operation, cannot be cost-free. Otherwise, the notion of "copy" loses its explanatory power. Instead, they proposed the pointer mechanism, where VPE only creates a link pointing to an existing memory representation, rather than duplicating the representation at the retrieval site. Processing the pointer is not affected by the complexity of the representation that the pointer links to.

However, the specific experimental stimuli used by Martin and McElree (2008) did not ensure that participants deeply processed the elliptical sentence. As discussed by Philips and Parker (2014) and Paape et al. (2017), the semantic compatibility of the experimental items entirely depended on the subject NP at the ellipsis site and the verb alone in the antecedent VP. The NP material included in the antecedent VP, which was designed to increase the antecedent complexity, was not necessarily used by participants for the purpose of making the semantic judgment. For instance, in example (5), in order to make a semantic compatibility judgment, readers only needed to notice that the "overly worn books" could never "understand" anything. Regardless of whether the object NP is simple ("Roman mythology") or complex ("Rome's swift and brutal destruction of Carthage"), only the "students" can be the subject of understanding something. Therefore, the semantic compatibility task was easily determined by the selection mismatch between the subject and the verb. The reader did not need to retrieve the full VP antecedent, including its object NP where the complexity manipulation was made, to make a correct compatibility judgment.

A recent study from Paape et al. (2017) used comprehension questions that could engage participants in more deeply processing the content of the VP antecedent. In addition, the study also added some follow-up regions after the ellipsis site "did too" to avoid the confound of sentence-final wrap-up effects, as well as adding more words between the antecedent and the VPE to avoid spillover from the antecedent. Half of the participants completed detailed comprehension questions as shown in (6), which target the content of the complex VP antecedent. The other half as a control group answered superficial questions that only asked if some phrase was mentioned in the sentence. In a self-paced reading experiment, participants faced with detailed questions would presumably process the VPE more deeply than participants in the superficial condition. Therefore, it is possible that a complexity effect could appear in the detailed probe condition while the superficial condition likely replicates the null effect of previous studies. However, their results found no effect on the ellipsis site "did too" from the complexity of the VP antecedent, regardless of the type of comprehension question probe.

(6) a. Simple: The advanced students loved *the afternoon session*, but as of late it was evident that the mathematics lecturer did not, as the time-consuming preparation really exhausted her.
b. Complex: The advanced students loved *the late afternoon session's many illustrative examples*, but as of late it was evident that the mathematics lecturer did not, as the time-consuming preparation really exhausted her.
c. Superficial probe: A mathematics lecturer was mentioned.
d. Detailed probe: A lecturer did not love an afternoon session's examples.

In this paper, we follow Paape et al.'s (2017) emphasis on engaging participants in deep processing of the VPE, and attempt to improve the use of the semantic compatibility judgment task building upon Martin and McElree's (2008) experiment. We believe the combination of these two design features helps better control the various methodological concerns discussed above. While the semantic compatibility judgment task in Martin and McElree (2008) successfully demanded the semantic integration of the subject and the verb, it fell short of ensuring that the participants indeed retrieved the whole VP antecedent instead of just the verb. Paape et al. (2017) focused on engaging the participants to fully retrieve the VP antecedent, but they did not take strong measures to encourage semantic integration. For instance, their detailed probe questions, although involved recalling the VP antecedent, did not necessarily prompt a full understanding of what the elliptical site actually

means given the VP antecedent. Therefore, we need an experiment task where the semantic judgment of an elliptical sentence requires a detailed probe into the VP antecedent. The complexity manipulation should be highly relevant to the judgments of the plausibility of the sentence. This improved control of deep processing produces a novel result that the ellipsis site "did too" is read faster when the VP antecedent is complex, but only when the sentence overall is implausible. We conclude that the representational complexity of the antecedent facilitates later processing at the VPE, but only when participants pay attention to the semantic implausibility of the sentence.

2 Self-Paced Reading Experiment

2.1 Design

This experiment adopts a 2x2 design (Complexity x Plausibility). Complexity is a within-subject variable with two levels: Simple vs Complex. Plausibility is also a within-subject variable with two levels: Plausible vs Implausible. Each trial consists of a self-paced reading of a sentence of one of the four conditions, followed by a binary plausibility judgment.

(7) An example stimuli item
   a. Simple Plausible: The doctor / handled / the / diagnosis. / The / expert / specialist / did too / in / the / busy / hospital.
   b. Simple Implausible: The doctor / handled / the / diagnosis. / The / young / salesperson / did too / in / the / busy / hospital.
   c. Complex Plausible: The doctor / skillfully / handled / the / tragically / incurable / diagnosis. / The / expert / specialist / did too / in / the / busy / hospital.
   d. Complex Implausible: The doctor / skillfully / handled / the / tragically / incurable / diagnosis. / The / trainee / nurse / did too / in / the / busy / hospital.

(7) shows an example item in four conditions. The complex VP antecedents (7c/d) were constructed by taking the simple VP antecedents in (7a/b) and adding one verb modifier ("skillfully") and two NP modifiers ("tragically incurable"). Slashes in (7) indicate chunks in the self-paced reading task. The implausible conditions (7b/d) were constructed by creating semantic incompatibility between the subject NP of the ellipsis clause and the full VP antecedent. In the Simple Implausible condition, it is unlikely that "the young salesperson" can "handle the diagnosis". In the complex implausible condition, a "trainee nurse" is unlikely to be able to "skillfully handle the tragically incurable diagnosis". However, the subject NP of the ellipsis clause was compatible with part of the VP antecedent, i.e. it is plausible for a "trainee nurse" to simply "handle the diagnosis". Therefore, detecting the implausibility in the complex VP condition requires the retrieval and integration of the entire VP antecedent, not just part of it. By asking participants to make a plausibility judgment, we hope to engage participants in a "deep processing" mode of the stimuli sentences, and fully retrieve and integrate the entire antecedent VP.

24 experimental items like those in (7) were constructed. An additional 24 filler sentences that did not contain VPE were also created, with 12 plausible fillers and 12 implausible fillers.

2.2 Plausibility Norming

The semantic plausibility manipulation in the experiment was verified in two norming experiments conducted on the PCIbex platform. In the first norming experiment, we removed the sentence-final PP region after "did too" (e.g. "in the busy hospital" in example (7)) and normed the plausibility of the sentence fragments up until the VPE site. This was done to ensure that the plausibility (or implausibility) of the sentences is determined by the VP ellipsis, rather than the sentence-final adjuncts. Forty-five native speakers of English between the ages of 18 and 50 completed the study on Prolific and were compensated $2.5. For each trial, a participant saw a sentence and rated the plausibility of the sentence on a 1-7 scale, 1 being most implausible and 7 being most plausible. Each participant rated 24 experimental sentences and 24 fillers.

	Plausible	Implausible
Simple	5.75 (sd=1.46)	2.34 (sd=1.47)
Complex	5.01 (sd=1.91)	3.05 (sd=1.67)

Table 1: Mean and standard deviation of plausibility ratings of experimental items without their final PP regions on a 1-7 scale (1 = most implausible; 7 = most plausible).

Table 1 shows the results of the norming experiment without the sentence-final PP regions. The plausibility manipulation was indeed effective (Est=1.96, se=0.16, $p<0.001$), as the mean plausibility ratings of plausible conditions are much higher than that of the implausible conditions. There is also a significant interaction between plausibility and complexity (Est=1.46, se=0.23, $p<0.001$). The complex plausible condition is rated less plausible than the simple plausible condition, and the complex implausible condition is rated less implausible than the simple implausible condition.

The second norming experiment was largely identical to the first one, except the sentence-final adjunct regions were added back to the norming sentences. Forty-eight participants completed the study on PCIbex (Zehr and Schwarz 2018). As shown in Table 2, the plausibility ratings on the whole-sentence stimuli are very similar to the first norming experiment, confirming that the (im)plausibility of the stimuli sentences is mainly driven by the ellipsis resolution, i.e. the integration of the subject NP of the ellipsis clause and the antecedent VP. The plausibility manipulation was effective (Est=1.49, se=0.17, $p<0.001$). The interaction between plausibility and complexity is also significant (Est=1.40, se=0.24, $p<0.001$), with the simple condition having the more strongly plausible or implausible ratings.

	Plausible	Implausible
Simple	5.49 (sd=1.56)	2.62 (sd=1.66)
Complex	5.03 (sd=1.69)	3.51 (sd=1.92)

Table 2: Mean and standard deviation of plausibility ratings of experimental items with their final PP regions on a 1-7 scale (1 = most implausible; 7 = most plausible).

2.3 Participant

41 native speakers of English between the ages of 18 and 50 participated in the study on Prolific. The experiment took about 15 minutes and each participant was compensated $4. Four participants were excluded from further analysis because they were less than 75% correct in the plausibility judgment of the filler trials.

2.4 Procedure

Participants performed a self-paced reading task implemented on the PCIbex platform (Zehr and Schwarz 2018), and read each sentence phrase by phrase (see example (7) for the segmentation of phrases). After each trial, they were presented with a prompt question "is the situation described by this sentence plausible or implausible?" Participants selected either "plausible" or "implausible" as their response. If their choice did not match the pre-defined plausibility label of each condition, they would receive an error message indicating they have made an error. The error message was designed to encourage participants to pay more attention to the materials. Each participant first completed four practice trials. The actual experiment contains 24 experimental trials and 24 filler trials randomly ordered.

2.5 Results

The critical region of interest is the ellipsis site ("did too"). We applied two statistical tests to the reading times. Firstly, we treated both plausibility and antecedent complexity as binary categorical variables and therefore analyzed reading times as results of a 2x2 design. Secondly, since the

norming experiment provides continuous plausibility ratings on a scale of 1-7, we treated plausibility as a continuous variable and each experimental sentence has a plausibility value between 1 and 7. Antecedent complexity was still treated as a categorical variable. In both cases, we conducted the Bayesian regression models on the data, including the two variables and their interactions as the fixed effects and the different participants and items as random effects. The models were implemented in R using the *brms* package (Bürkner 2017). Outlier reading times over 2000ms were excluded from the analysis, and RTs were log-transformed prior to the analysis. In all the models, the priors for all the parameters were set to be Normal(0, 10).

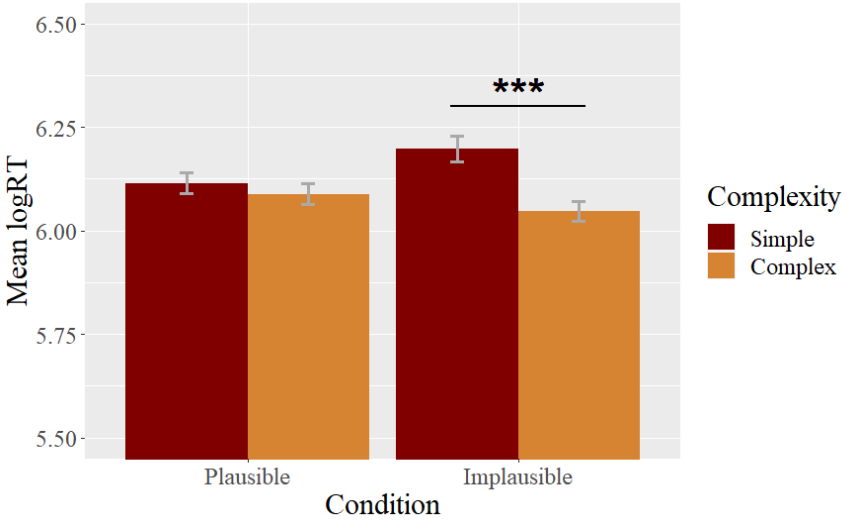

Figure 1: Mean log RT at the critical region by condition.

The results with a categorical Plausibility variable is shown in Figure 1. There is an effect of Complexity (Est = -0.06, se = 0.02, 95%CI = [-0.10, -0.02]), with the complex condition read faster than the Simple conditions, but this should be interpreted with caution since there is also evidence for a Complexity x Plausibility interaction (Est = -0.12, se = 0.04, 95%CI = [-0.20, -0.03]). The complex condition was read faster than the simple condition only when the sentences were implausible (p=0.002), but there was no complexity effect in the plausible conditions (p>.6).

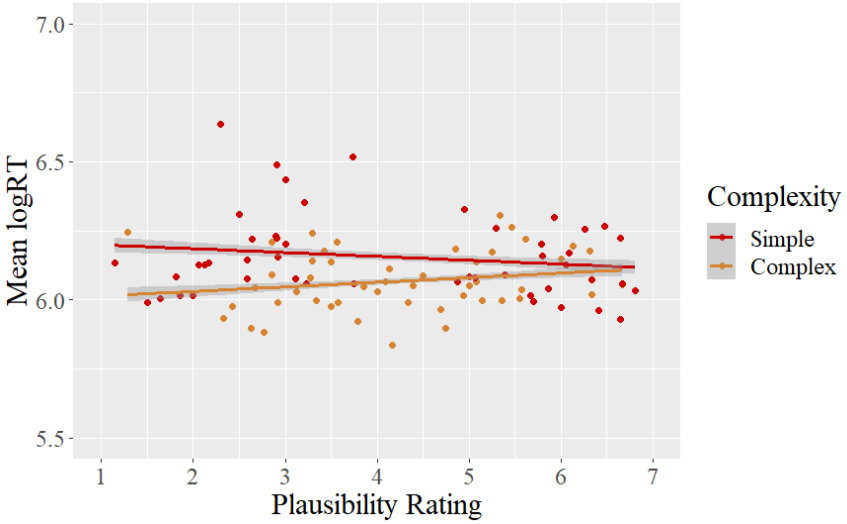

Figure 2: Mean log RT at the critical region by continuous plausibility ratings.

Figure 2 shows the analysis with a continuous plausibility rating variable. Towards the lower end of the plausibility scale (indicating higher implausibility), the complex sentences had a faster reading time than the simple sentences. As the plausibility ratings go up, the reading time gradually converges between the simple and the complex sentences. This result is entirely in line with the first analysis above. The Bayesian model shows evidence for a Complexity effect (Est = 0.19, se = 0.06, 95%CI = [0.08, 0.30]) and a Complexity x Plausibility interaction (Est = -0.03, se = 0.01, 95%CI = [-0.05, 0]).

3 Discussion

After carefully controlling for our experimental material and also engaging the participants in a deep processing mode, our study was able to demonstrate an antecedent complexity effect in ellipsis resolution. Interestingly, however, this complexity effect only appeared for implausible sentences: complex implausible sentences were read faster than simple ones at the ellipsis site, but there was no difference for simple vs complex plausible sentences. The plausibility-modulated complexity effect is not predicted by any of the existing accounts. It has several implications for ellipsis resolution and more generally also the memory retrieval mechanism for non-local dependencies. We discuss these implications below.

Firstly, the results cast some doubts on the copy mechanism proposed by Murphy (1985), which predicts that copying a more complex antecedent relative to a simpler one should be more costly and hence take a longer time. In our experiment, the complexity of the VP antecedent does not increase the reading time of the sentences in either the plausible or the implausible condition. The null effect of antecedent complexity in the plausible conditions seems to be more in line with the predictions of a pointer mechanism (Frazier and Clifton 2001, Martin and McElree 2008, Paape et al. 2017). This mechanism does not assume any operation of "copying" the antecedent, nor does it assume additional cost of retrieving and integrating more complex antecedent with the subject NP at the ellipsis site. The drawback, however, is that this account in its current form does not immediately account for the complexity effect in the implausible conditions.

Beyond ellipsis processing, our findings also have implications on the memory retrieval system in human sentence processing. Several previous studies have found that, for relative clause processing and pronoun resolution at least, increasing the representational complexity of the target facilitates its later retrieval (Hofmeister 2011, Hofmeister and Vasishth 2014, Troyer et al. 2016, Karimi et al. 2018, 2020). There could be a number of explanations for the complexity facilitation effect. A retrieval target that is more complex is also generally more semantically elaborated. This could increase the salience of the target representation in working memory, facilitating later retrieval (Hofmeister 2011). The longer time required to encode a complex target is likely to result in more attention allocated to the target representation, also facilitating its retrieval later (Karimi et al. 2020). Due to its enhanced distinctiveness, a more complex memory representation with elaborated properties also reduces the likelihood of similarity-based interference with other memory representations, again facilitating later retrieval (Lam and Xiang 2022). These are not mutually exclusive possibilities.

In the current study, the facilitation effect of complexity in our implausible conditions appears to be consistent with previous results. Specifically, elaborated features in the VP antecedent may have facilitated the detection of semantic incompatibilities, resulting in the complexity facilitation effect in the implausible conditions. For instance, in example (7d), additional features introduced by the modifier "tragically incurable" provide more cues to help participants detect the semantic incompatibility between "the trainee nurse" and the VP "handled the tragically incurable diagnosis". In contrast to this, in the simpler implausible condition (7b), there may be fewer cues for people to quickly reach the conclusion that a "young salesperson" is unlikely to "handle the diagnosis". We also note that some previous studies have found that an implausible antecedent can slow down retrieval. Hofmeister (2011) compared the processing of relative clauses where the head noun of the RC was either a typical NP (e.g. "the ruthless military dictator) or an atypical NP (e.g. "a lovable military dictator") – in other words, plausible or implausible. The results showed that atypicality increased reading times during the encoding of the head noun, indicating more encoding efforts, but it also increased retrieval difficulty at a later retrieval site. The major difference between Hofmeister's manipulation and the current experiment is that the implausibility in Hofmeister's materials

arises due to the retrieval target itself, while the implausibility of our materials is due to the incompatibility between the subject NP at the ellipsis site and the VP antecedent, but the antecedent VP itself is coherent.

In the meantime, however, the null effect of antecedent complexity in our plausible conditions is inconsistent with previous findings. We suggest a few possibilities, all of which may lend particular support for the last mechanism outlined above, i.e. facilitation of memory retrieval due to reduced interference between memory representations. We note an important difference between the current study and previous studies that observed complexity facilitation effects in relative clauses and pronoun resolution. While the retrieval target in the current study was a VP antecedent, the retrieval targets in previous studies were NPs. This difference poses interesting questions about theories of similarity-based interference. In previous studies on relative clauses and pronoun resolution, in addition to the target NP, there were also other NPs in the same sentence, resulting in interference among their memory representations. The more distinct the target NP was, the smaller the probability for similarity-based interference to arise, which explains why a complex target NP could facilitate its later retrieval. In the current study, however, when memory retrieval takes place at the ellipsis site, there are no additional VPs in working memory other than the antecedent target. We therefore did not observe any difference between a complex vs. a simple target VP, since similarity-based interference was not an issue here. Related to this, previous work on similarity-based interference primarily focused on interference between NPs, and it is an open question whether and how interferences arise due to competition between VP representations, in situations when there are multiple VPs in the same sentence. More future work is needed to address this question.

In conclusion, by adopting an experimental design that encourages detailed processing of VPE resolution, we made a novel empirical observation that the representational complexity of a VP antecedent facilitates ellipsis resolution, but the facilitation effect is constrained by semantic plausibility. Our findings have several implications for both theories of VPE resolution and theories of memory retrieval mechanisms.

References

Bürkner, Paul-Christian. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80:1–28.

Frazier, Lyn, and Charles Clifton. 2000. On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research* 29(2):125–140.

Frazier, Lyn, and Charles Clifton. 2001. Parsing coordinates and ellipsis: Copy α. *Syntax* 4:1-22.

Hofmeister, Philip. 2011. Representational complexity and memory retrieval in language comprehension. *Language and cognitive processes* 26(3):376–405.

Hofmeister, Philip., and Shravan Vasishth. 2014. Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in psychology* 5:1–13.

Just, Marcel A., and Patricia A. Carpenter. 1980. A theory of reading: from eye fixations to comprehension. *Psychological review* 87:329–354.

Karimi, Hossein, Tamara Y. Swaab, and Fernanda Ferreira. 2018. Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *Journal of Memory and Language* 102:68–82.

Karimi, Hossein, Michelle T. Diaz, and Eva Wittenberg. 2020. Sheer Time Spent Expecting or Maintaining a Representation Facilitates Subsequent Retrieval during Sentence Processing. In *CogSci.* 2020:2728–2734.

Lam, Chi Dat Daniel, and Ming Xiang. 2022. Noun phrase representational complexity reduces maintenance cost in working memory by increasing distinctiveness between referents. In *Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 44)*, ed. J. Culbertson, A. Perfors, H. Rabagliati and V. Ramenzoni, 1298–1305.

Martin, Andrea E., and Brian McElree. 2008. A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language* 58:879–906.

Murphy, Gregory L. 1985. Processes of understanding anaphora. *Journal of Memory and Language* 24(3):290–303.

Paape, Dario, Bruno Nicenboim, and Shravan Vasishth. 2017. Does antecedent complexity affect ellipsis processing? An empirical investigation. *Glossa: a journal of general linguistics* 2:77.

Phillips, Colin, and Dan Parker. 2014. The psycholinguistics of ellipsis. *Lingua* 151:78–95.

Tanenhaus, Michael K., and Greg N. Carlson. 1990. Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes* 5:257–280.

Troyer, Melissa, Philip Hofmeister, and Marta Kutas. 2016. Elaboration over a discourse facilitates retrieval in sentence processing. *Frontiers in psychology* 7:374.
Zehr, Jeremy, and Florian Schwarz. 2018. PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832

Department of Linguistics
The University of Chicago
Chicago, IL 60637
danielyue@uchicago.edu
mxiang@uchicago.edu