Covariation in processing: Grammar vs. context

Nikhil V. Lakhani, University of Pennsylvania, US, lakhanin@sas.upenn.edu

Florian Schwarz, University of Pennsylvania, US, florians@sas.upenn.edu

In addition to referential uses, pronouns can have covarying interpretations, i.e., exhibit the behavior of a bound variable. The grammatical mechanism(s) behind such readings have been subject to longstanding debates: Some authors argue for a fairly flexible but unified semantic mechanism that is not tied closely to syntactic configurations. Others distinguish a core class of *bona fide* binding with tight syntactic constraints from other mechanisms that give rise to ultimately parallel effects, but do so more indirectly. Psycholinguistic work has started to uncover the processing mechanisms involved in evaluating dependencies between covarying pronouns and (candidate) antecedents. Moulton and Han (2018) leverage the processing perspective to try to shed light on the theoretical question of what grammatical mechanism is at play for a given covarying pronoun. They argue that so-called Gender Mismatch Effects only arise for cases of *bona fide* binding, supporting the existence of distinct grammatical mechanisms. However, Kush and Eik (2019), looking at another construction involving the relevant other covariation mechanisms, do find Gender Mismatch Effects. These authors suggest that various contextual factors can make a covarying interpretation harder to obtain, and they propose adjustments to Moulton and Han's stimuli that they think should lead to fast Gender Mismatch Effects even when no *bona fide* binding is involved. A series of self-paced reading experiments replicate the results from Moulton and Han, and then extend the paradigm to variations along the lines suggested by Kush and Eik. The adjustment of contextual factors indeed results in Gender Mismatch Effects for both environments. We discuss how the processing evidence informs the theoretical issues.

1. Introduction

1.1 Requirements for covariation: C-command vs. semantic scope

In addition to referential uses, pronouns can have covarying interpretations, i.e., exhibit the behavior of a bound variable. According to the classic account by Reinhart (1983), binding requires the antecedent to *c-command* the pronoun.[1] This view captures contrasts such as the following (adapted from Barker, 2012):

(1) a. [$_{DP}$ **The woman** who traveled with the man]$_i$ denied that **she**$_i$ met the shah.
     b. [$_{DP}$ **Each woman** who traveled with the man]$_i$ denied that **she**$_i$ met the shah.

(2) a. The [man who traveled with [$_{DP}$ **the woman**]$_i$] denied that **she**$_i$ met the shah.
     b. # The [man who traveled with [$_{DP}$ **each woman**]$_i$] denied that **she**$_i$ met the shah.

In (1), the noun phrase *the woman who traveled with the man* c-commands the pronoun *she* and can serve as its antecedent, both when it is referential (1a) and quantificational (1b). In (2), the antecedent noun phrase *the woman* occurs inside of the relative clause within the subject, thus not c-commanding the pronoun *she* in the verb phrase. Here, a *coreferential* interpretation of the pronoun with the referential DP is still possible (2a), as coreference does not require c-command. But the minimal variant with a quantifier phrase antecedent in (2b) does not allow a *covarying* reading, as binding is unavailable, due to the lack of c-command. For ease of reference and consistency with prior work, we'll refer to quantifier phrases as QPs, but this should not be taken to suggest a syntactic category distinct from DP. However, Barker (2012) catalogues a variety of both previously known and novel exceptions to the c-command requirement. For example, non-c-commanding QPs in possessive DPs (3a), inverse linking constructions (3b), PPs (3c), and VPs (3d) from Barker can serve as antecedents for a pronoun receiving a covarying interpretation:

(3) a. [No one$_i$'s mother-in-law] fully approves of her$_i$.
     b. [Someone from every city$_i$] hates it$_i$.
     c. John gave [to each participant$_i$] a framed picture of his$_i$ mother.
     d. John [visited each student$_i$] on his$_i$ birthday.

Various types of reactions to considerations of this type have been offered in the literature, which fall into two broad camps: Unified accounts of covarying pronoun interpretations pursue a reconceptualization of the notions of binding and covarying interpretations that covers all of these cases. In contrast, two-mechanism accounts propose an alternative grammatical mechanism for covarying interpretations of pronouns with non-c-commanding quantificational antecedents, distinct from *bona fide* binding, with c-command only required for the latter.

---

[1] While precise definitions of c-command vary in the literature, they are generally based on the notion that for X to c-command Y, Y has to be properly contained in the syntactic sister node of X.

Barker (2012) represents a prominent instance of a unified account (building on Safir, 2004; see Barker's discussion for a review of alternative options, including adjustments to the definition of c-command). The account replaces the c-command requirement for binding with a weaker scope requirement: A QP can bind a pronoun if it can take scope over the position of the pronoun. This still excludes (2b), because the QP cannot take scope over that position. This is shown in (4), where another scope-taking expression (*a student*) appears in place of the pronoun in (2b), which the QP cannot scope over.

(4)   The [man who traveled with each woman] denied that a student met the shah.
        CANNOT MEAN: *For each woman x, it holds that the man y who traveled with x denied that there is some student z that met the shah.*   *(each woman > a student)*

But the scope requirement allows (3a–d), since the relevant QP can take scope over the position of the pronoun there. This is shown (for (3b) and (3d)) by the possible interpretations in (5), where the QP takes scope over the indefinite replacing the pronoun:

(5)   a.   [Someone from every city] hates a city regulation.
             CAN MEAN: *For every city x there is some person y, such that there exists a city regulation z and y hates z.*   (*every city > a city regulation*)
        b.   John [visited each student] on a Monday.
             CAN MEAN: *For each student x there is some Monday y, such that John visited x on y.*
                                                                    (*each student > a Monday*)

Different two-mechanism accounts have proposed various mechanisms for letting a pronoun covary with a QP in the absence of c-command and binding. The most popular mechanism involves quantification over situations (see, e.g., Büring, 2004, for a situation-based account of (3a)). It is not clear whether quantification over situations can be used across all cases of covariation without c-command, but the mechanism can be employed with the types of sentences we consider in our experiments. We spell out a situation-based analysis of our experimental sentences in Section 5.

1.2 Previous experimental work on covarying pronouns

Several lines of recent psycholinguistic work on the processing of covarying pronouns have investigated the role of c-command in online processing, using so-called *Gender Mismatch Effects* (GMMEs; Sturt, 2003): Relative to cases where a pronoun is anteceded by a noun phrase of matching gender, reading times on the pronoun and following region(s) increase when there's a mismatch in gender between the two. This is taken to indicate a disruption of some sort caused by the gender mismatch.[2] Sturt's original study used this approach to test reflexives

---

[2] The precise nature of this disruption can, in principle, be construed in various ways. Perhaps the most prominent view, adopted by cue-based retrieval approaches (discussed in 5.2), is that the mismatching antecedent is not

for effects of Principle A of the binding theory. GMMEs, then, provide a useful diagnostic for whether the processor is attempting to establish a dependency between a given pronoun and a potential antecedent. An absence of such an effect suggests the relevant potential antecedent is not considered at all. Cunnings et al. (2015) and Kush et al. (2015) use GMMEs to test if c-command limits access to QPs as potential antecedents during pronoun processing, or whether non-c-commanding QPs can be considered. Cunnings et al.'s findings suggest that c-command affects whether a QP can be considered. The authors found GMMEs in sentences like (6b) (where *every old man* c-commands the pronoun: *CC*), but not (6a) (where it doesn't: *NoCC*).

(6) a. The surgeon who every old man on the emergency ward saw silently wished that {(i) he / (ii) she} could go a little bit faster. (*NoCC* (i) *Match* / (ii) *Mismatch*)

  b. The surgeon saw that every old man on the emergency ward silently wished that {(i) he / (ii) she} could go a little bit faster. (*CC* (i) *Match* / (ii) *Mismatch*)

Further experiments by Kush et al. (2015) corroborate these findings. Taken together, these results suggest that the processor does not attempt to establish a dependency between QP antecedents and pronouns when the former do not c-command the latter. More generally, they show that relations between items, rather than item-specific information such as morphological features alone, are relevant for the retrieval of the QP antecedent. We discuss more details of Kush et al.'s cue-based approach, which incorporates these results, in 5.2. Empirically, their results add to a growing body of research on the impact of other structural constraints in online antecedent retrieval (e.g., Sturt, 2003, on Principle A; Chow et al., 2014, on Principle B).

However, Cunnings et al. (2015) and Kush et al. (2015) do not specifically tease out c-command as the relevant structural constraint at play. Given that their primary interest was in testing for relational constraints in general, these authors intentionally chose constructions where c-command and Barker's (2012) scope constraint make the same predictions: A QP in a relative clause cannot take scope over the main clause verb phrase in (6). This leaves open the possibility that what governs the accessibility of the antecedent is a matter of scope, not c-command. In an attempt to differentiate between these two possibilities, Moulton and Han (2018) aim to garner evidence for a two-mechanism approach to covariation, where c-command does have a special role to play. Their Experiment 2 presents a variant of the general Cunnings et al. experiment, using stimuli with a QP in sentence-initial temporal adjunct clauses:

considered for pronoun resolution. The observed delays are then attributed to difficulties in interpreting the pronoun relative to a suitable antecedent, whether it is because there isn't one at all or because it is less easily accessible than in sentences with the gender-matching variant. Alternatively, one could posit that the mismatching antecedent is temporarily considered for interpreting the pronoun, which then leads to a clash based on the feature mismatch. Our discussion is largely independent of this, but see note 6 and the discussion of Experiment 2 for some further relevant considerations.

(7) a. After each boy brought fresh water from the kitchen quickly it seems that {(i) he / (ii) she} went on an early break. (*QP* & *NoCC* (i) *Match* / (ii) *Mismatch*)

b. It seems each boy brought fresh water from the kitchen quickly right before {(i) he / (ii) she} went on an early break. (*CC* (i) *Match* / (ii) *Mismatch*)

Crucially, and unlike in Cunnings et al. (2015), the no-c-command (NoCC) condition here allows the quantifier to take scope over the pronoun in the absence of c-command, as confirmed in an offline judgement task. This manipulation thus makes it possible to distinguish the processor's sensitivity to scope and c-command, unlike Cunnings et al. and Kush et al. (2015): If scope is decisive, GMMEs are expected in both conditions. If c-command is what matters, we'd only expect a GMME in the CC condition. In a self-paced reading (SPR) task, Moulton and Han (2018) find an initial interaction indicative of GMMEs in the CC condition. but not the NoCC condition. Following the general GMME logic, they interpret this as showing that the processor was not attempting to establish a dependency between the pronoun and the potential QP antecedent in the NoCC condition, in contrast to the CC condition. A second SPR experiment by Moulton and Han (Experiment 3) compares sentences with exceptionally covarying pronouns with QP antecedents, like (7a), to identically structured sentences with DP antecedents, as in (8). This yielded a parallel interaction, due to the presence of a GMME in the DP condition, where the referential interpretation is not dependent on c-command, and the absence thereof in the QP condition.

(8) After the boy brought fresh water from the kitchen quickly it seems that {(i) he / (ii) she} went on an early break. (*DP* (i) *Match* / (ii) *Mismatch*)

In theoretical terms, Moulton and Han (2018) argue that their results support a two-grammatical-mechanism view: While covariation involving standard binding requires c-command, the mechanism at play in other cases, such as (7a), does not. With regards to processing, the question is why this mechanism does not give rise to GMMEs. Moulton (2017) spells out a specific grammatical proposal that aims to answer this. On this view, exceptionally covarying pronouns are interpreted as D-Type pronouns (Elbourne, 2005; Postal, 1966), which, underlyingly, are definite descriptions with a phonologically null noun phrase. The interpretation of the D-Type pronoun is relativized to a situation variable. Quantification over this situation variable, rather than a standard variable over individuals, is what yields a covarying interpretation of the D-Type pronoun (adapting the analysis of temporal adjunct clauses in Artstein, 2005; see 5.3 for more details). Illustrating informally for (7a), for any given situation s containing a boy who brought water, there is a temporally later situation in which the unique boy in s went on a break. Note that, formally, only the situation pronoun, which has no gender features, is bound and quantified over. The pronoun *he* and the candidate

antecedent do not stand in a formal binding relation. The crucial processing repercussion of this is that there is no evaluation of whether or not their gender features match, at least not initially. The interaction found by Moulton and Han, based on the absence of GMMEs in the NoCC condition, is then explained as follows: Feature match only plays a role in (initial) processing in standard binding configurations, where the individual variable denoted by the pronoun is quantified over. It does not play a role in the processing of covarying pronouns in non-c-commanding configurations which involve quantification over situation variables.

Note that the semantic misalignment that comes with the gender mismatch ultimately matters in non-c-commanding cases, too, of course, since the Mismatch version of (7a) does not allow for a covarying reading. But, the story must go, this doesn't happen until some later point, and it may not give rise to the same processing effect as gender mismatch for a pronoun whose individual variable is potentially bound under c-command even then, depending on how the nature of GMMEs is construed. However, if one construes delays for mismatching antecedents as involving an initial attempt at forming an anaphoric dependency which then falters in light of the incompatible gender features (see note 2), then perhaps ultimately Moulton and Han's (2018) account predicts later effects that are parallel in nature, as the feature mismatch doesn't arise in the initial phase of considering pronoun and antecedent, when only the situation pronoun and its abstract binder are considered.

A straightforward prediction of this proposal, based on a situation semantic account of covariation, is that other cases subject to such an analysis should behave similarly. Two studies have looked at related issues: Earlier work by Carminati et al. (2002) compared so-called telescoping sentences without c-command, such as *Every British soldier aimed and then he killed an enemy soldier,* with parallel variants allowing for binding under c-command. Comparing DP and QP antecedents (but not testing for GMMEs), they found no processing costs associated with a non-c-commanding QP antecedent. This suggests that the relevant two types of covariation mechanisms do not necessarily differ in their processing time-course, but this is consistent with Moulton and Han's (2018) proposal, since GMMEs are not at play. Kush and Eik (2019), however, directly test the relevant prediction by looking at donkey pronouns (9a), one of the most prominent cases for which situation semantic D-Type analyses have been proposed (Elbourne, 2005; Heim, 1990):

(9)   English paraphrases of Kush and Eik's (2019) Norwegian sample stimuli:
        a.   Every father who had a daughter in a soccer league drove {her / him} to the games.
        b.   The father who had a daughter in a soccer league drove {her / him} to the games.

Using referential pronouns (9b) for a DP-baseline comparison, Kush and Eik (2019) conducted an SPR study in Norwegian, in a standard GMME design with gender matching or mismatching pronouns. They find GMMEs in both cases, and no interaction, suggesting that the non-c-

commanding indefinite antecedent and its particular gender features were accessible in early processing for both referential and covarying interpretations. Assuming a situation semantic analysis of donkey sentences, this directly contradicts the key prediction from Moulton and Han (2018), spelled out above.

The contrast in GMME findings for the NoCC stimuli in Moulton and Han (2018) and Kush and Eik (2019) raises a new question about the sources of GMMEs under covariation. In theoretical terms, there certainly are differences between the constructions in play that could have repercussions for processing. While the cases with QPs in temporal adjuncts, as in (7a), are among those that are standardly considered to fall under the scope constraint according to Barker (2012), donkey sentences are not (and neither are cases of telescoping).[3] This is because they do not involve the antecedent QP taking scope over the position of the pronoun. In contrast, QPs in temporal adjuncts can take scope over the position of the pronoun. Just why this difference should lead to the pattern of GMMEs reviewed here remains open at this point, but it's worth noting these differences of potential relevance. Kush and Eik suggest an alternative processing proposal. Their account maintains that a single processing mechanism uniformly governs antecedent retrieval in covarying cases (as well as referential ones), with or without c-command. But it allows for possible variation between specific types of cases. In particular, they suggest that the accessibility of the antecedent is affected not just by structural factors like c-command, but also by contextual factors at play in settling on an overall interpretation, specifically with regards to scope. The idea that such factors can have an impact is not new, and has been discussed for telescoping in some detail (Anderssen, 2011; Poesio & Zucchi, 1992). Kush and Eik suggest a number of specific changes that they speculate should increase the availability of covarying readings in the Moulton and Han type stimuli, and thus lead to more immediate GMMEs.

1.3 The present approach

We report a series of self-paced reading experiments to further inform these debates. Our first goal is to answer the empirical question: How does the processing profile of exceptionally covarying pronouns differ from ordinary bound pronouns? And, to the extent that it does, is this modulated by contextual factors? We first report a replication of Moulton and Han's (2018) original study, to ensure that we are starting from the same baseline. We then implement adjustments of contextual factors suggested by Kush and Eik (2019), to assess what effect, if any, this has on the presence of GMMEs. As sketched above, in order to explain the absence

---

[3] Though this ultimately depends on your analysis of these sentences, and the details of how scope is construed in a given semantic framework. Note that Barker and Shan (2008) analyze donkey pronouns as involving in-scope binding in continuation semantics.

of GMMEs in their exceptionally covarying sentences, Moulton and Han advance an account that implies immediate accessibility of the QP antecedent in all cases of covariation, but ties GMMEs to purely structural factors, i.e., c-command. In this account, non-structural contextual or content variations within the same exceptionally covarying configuration would play no role in the presence of GMMEs. In contrast, Kush and Eik's take on the Moulton and Han results is that the relevant contextual pressures for a covarying interpretation of the stimuli in Moulton and Han (2018) were not strong enough to make the QP immediately accessible. Rather, participants landed on covarying interpretations at later stages in the comprehension process. According to this account, the presence of GMMEs is affected by non-structural factors, such as the availability of a given scope interpretation, and appropriate manipulations to the stimuli should reflect this.

Our second goal is to assess the theoretical implications of the overall body of empirical evidence. If we fail to find GMMEs in the exceptionally covarying condition, even after contextual adjustments, this would further support the argument in favor of two-grammatical-mechanisms accounts put forward by Moulton and Han (2018). We can easily attribute processing differences between cases of covariation to the inherent theoretical contrast these accounts posit. On the contrary, a unified grammatical perspective does not provide an inherent explanation of such processing differences. Importantly, the converse does not hold: The presence of GMMEs in exceptionally covarying cases does not necessarily, in and of itself, argue against two-mechanism grammatical accounts. This is because different theoretical derivations of an interpretation need not correspond to processing differences. But the lack of such processing differences would take away Moulton and Han's specific line of argument in favor of two-mechanism accounts.

To preview, our results show that the adjustment of contextual factors indeed results in GMMEs with parallel time-courses for both environments.

2. Experiment 1

Our first experiment is a replication of Experiment 3 from Moulton and Han (2018). The critical experimental stimuli were identical to the original ones, crossing antecedent type (QP vs. DP) with gender match vs. mismatch, as illustrated in (10). However, we were only able to include 20 of the original study's 36 filler sentences. The replication ensures that this and other potential minor deviations in methods did not affect the GMME pattern, to provide a sound comparison with the results for the modified stimuli in Experiment 2.

(10)　　a.　After $_1/_2$ each boy $_2/_3$ brought fresh water $_3/_4$ from the kitchen $_4/_5$ quickly $_5/_6$ it seems $_6/_7$ that he $_7/_8$ went $_8/_9$ on an early $_9/_{10}$ break. *(QP Match)*

　　　　b.　After $_1/_2$ each boy $_2/_3$ brought fresh water $_3/_4$ from the kitchen $_4/_5$ quickly $_5/_6$ it seems $_6/_7$ that she $_7/_8$ went $_8/_9$ on an early $_9/_{10}$ break. *(QP Mismatch)*

c.   After $_1/_2$ the boy $_2/_3$ brought fresh water $_3/_4$ from the kitchen $_4/_5$ quickly $_5/_6$ it seems $_6/_7$ that he $_7/_8$ went $_8/_9$ on an early $_9/_{10}$ break. *(DP Match)*

d.   After $_1/_2$ the boy $_2/_3$ brought fresh water $_3/_4$ from the kitchen $_4/_5$ quickly $_5/_6$ it seems $_6/_7$ that she $_7/_8$ went $_8/_9$ on an early $_9/_{10}$ break. *(DP Mismatch)*

As a reminder, a replication of Moulton and Han's (2018) results crucially should involve an interaction between antecedent type and gender match in the critical and/or spillover regions, with GMMEs only arising in the DP conditions.

2.1 Materials and procedure

The 20 critical items from Experiment 3 in Moulton and Han (2018), varying by the four conditions in (10), were distributed across four lists in a Latin Square design, with individual participants only seeing each item in one condition. In addition, 20 of the 36 fillers from Moulton and Han, with structures unrelated to the manipulation of interest, were included, so that participants saw a total of 40 items. These were presented in randomized order, with an alternating pattern of a critical item followed by a filler. Each sentence was split into ten regions in a moving-window self-paced reading paradigm; starting from a set of dashes replacing each character on the screen, participants advanced to the next region by pressing the space bar (previous regions turned back into dashes). Each trial was followed by a yes/no comprehension question. The questions asked about the content of the sentences, but they were orthogonal to the manipulation, so as to not interfere with the data. The study was hosted on Ibex, an online experiment hosting service, and took place remotely. Participants were instructed to be in a quiet place without distraction. They received explicit instructions on doing the task, as well as three practice trials with feedback on the accuracy of their answers to comprehension questions. The total experiment lasted about 10 minutes. A link to a copy of the experiment can be found in (20) in Appendix A.

2.2 Participants

Eighty-three undergraduate students, who self-identified as native speakers of English, were recruited through the University of Pennsylvania's subject pool and received course credit for their participation. In line with subject pool policy, they saw a debriefing about the main research questions addressed by the experiment at the end.

2.3 Analysis

Prior to statistical analysis, data from participants with an accuracy rate below 70% on the comprehension questions (across all conditions and fillers) or an average reading time (RT)

below 300 ms (across all regions) were removed.[4] This eliminated two participants, leaving data from 81 participants to analyze. Individual trials were also removed if any one region's RT during that trial was above 3000 ms; this eliminated 60 experimental trials (4%).

Statistical analyses used the natural log-transformed RTs as the dependent variable (as in Moulton & Han, 2018). For each region, a linear mixed-effects model analysis using the lme4 package (Bates et al., 2015) was conducted in R (version 4.2.2). For this initial replication, we report two analysis variants: First, one analysis including only the manipulated factors as predictors, as reported by Moulton and Han, so as to have a fully parallel point of reference. Second, since it has become standard practice to include reading times for the previous region as an additional predictor in analyses of self-paced reading data, we also report an analysis with the natural log-transformed RT of the previous region added in as a fixed effect.[5]

The first model included fixed effects of antecedent type, gender match, and their interaction. Antecedent type and gender match were sum-coded, with one level as –1 and the other as 1. The initial model used a maximal random effects structure (Barr et al., 2013), with random intercepts and random slopes and interactions for participants and items. In case of convergence issues or random effects correlation issues, the random effects structure was gradually simplified by removing individual random effects slopes and by removing correlations with random slopes. For region 7, this resulted in a model with only random intercepts; for region 8, a model with (uncorrelated) by-participant slopes for antecedent type and gender match and a by-item slope for gender match; and for region 9, a by-participant slope for antecedent type and a random intercept for items. Planned comparisons to measure GMMEs per antecedent type were computed using the emmeans package. *P*-values were determined using the lmerTest package via the Satterthwaite method (Kuznetsova et al., 2017).

2.4 Results

Mean accuracy on the comprehension questions after data removal across all conditions and fillers was 0.90 (*SE* = .008). **Table 1** shows the mean accuracy rate broken down by condition, which shows no major effect on comprehension.

---

[4] In their Experiment 2, Moulton and Han (2018) report removing two participants with an average RT below 400 ms, who also had the two lowest comprehension rates (mean 62%); after removal, the lowest comprehension rate was 75%. In their Experiment 3, they removed no participants at all, and the lowest comprehension rate was 70%. We adjusted the RT-based cut-off to averages lower than 300 ms, as our participants seem to have been somewhat faster, though there also is no straightforward comparison, since we were missing 16 of Moulton and Han's fillers. Generally, even our faster readers had high comprehension question accuracy rates, suggesting sufficient engagement with, and comprehension of, the stimuli.

[5] Thanks to the handling editor, Ming Xiang, for suggesting we include the latter analysis.

Table 1: Mean accuracy rates of comprehension question responses (*SE*) in Experiment 1.

	Match	**Mismatch**
QP	.89 (.022)	.89 (.018)
DP	.92 (.016)	.90 (.018)

The graphs in **Figure 1** show natural log-transformed mean RTs for each region after data removal. Region 7 is the critical region containing the pronoun. Regions 8 and 9 are considered spillover regions, as effects from the manipulation in region 7 may emerge here as well.
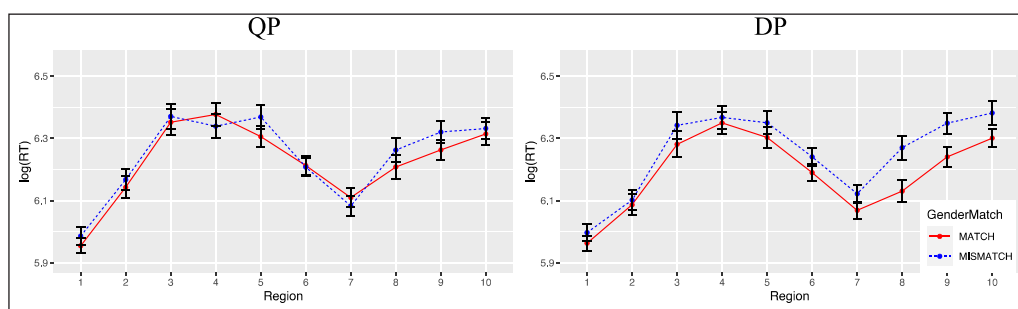
Figure 1: Log-transformed mean RTs (with standard errors) by region in Experiment 1.[6]

An overview of the interaction analyses can be found in **Table 2**, and results by region are summarized below.

Table 2: Summary of statistical analysis for Experiment 1 (without previous region RT as a predictor).

	Region 7 (pronoun)			**Region 8 (spillover)**			**Region 9 (spillover)**		
	Est.	*SE*	*t*	**Est.**	*SE*	*t*	**Est.**	*SE*	*t*
Antecedent Type	0.000	0.007	–0.022	–0.018	0.009	–2.028*	0.003	0.009	0.379
Gender Match	–0.008	0.007	–1.084	–0.047	0.013	–3.539**	–0.042	0.008	–4.959***
Type × Match	–0.020	0.007	–2.726**	–0.020	0.009	–2.255*	–0.009	0.008	–1.077

. $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

---

[6] In the DP condition, it can be seen on the graph that the gender mismatch condition is consistently slightly higher than the gender match condition, even before the critical region 7. This can be safely ignored, as the stimuli were identical up to that point, and further analysis showed no significant interactions ($p > 0.05$) until the critical region.

Region 7: The analysis revealed a significant interaction of antecedent type and gender match. Planned comparisons found a significant simple effect of gender match in the DP condition (Est. = –0.056, *SE* = 0.021, *t* = –2.700, *p* < 0.01), but not in the QP condition (Est. = 0.024, *SE* = 0.021, *t* = 1.159, *p* = 0.247).

Region 8: The analysis revealed a significant main effect of antecedent type as well as of gender match, and a significant interaction of antecedent type and gender match. Planned comparisons found a significant simple effect of gender match in the DP condition (Est. = –0.133, *SE* = 0.032, *t* = –4.200, *p* < 0.001) and a marginally significant one in the QP condition (Est. = –0.054, *SE* = 0. 032, *t* = –1.710, *p* = 0.093).

Region 9: The analysis revealed a significant main effect of gender match. Planned comparisons found a significant simple effect of gender match in both the DP condition (Est. = –0.102, *SE* = 0.024, *t* = –4.274, *p* < 0.001) and the QP condition (Est. = –0.066, *SE* = 0.024, *t* = –2.731, *p* < 0.01).

As we'll discuss in more detail below, these results, based on analyses parallel to those reported in Moulton and Han (2018), overall align with those in the original paper, confirming the comparability of our methods and data. But before discussing the interpretation of the data in more detail, let us offer a second set of analyses, which add the reading time in the preceding region as a further factor to the various models. (The random effects structures here were the same as above, with the exception of having to remove the by-participant random effects correlation for region 9.) The results are summarized in **Table 3**. Unsurprisingly, the reading time of the previous region was a highly significant predictor, but we will not comment further below on this, given that it's not of direct theoretical interest.

Table 3: Summary of statistical analysis for Experiment 1 (with previous region RT as a predictor).

	Region 7 (pronoun)			Region 8 (spillover)			Region 9 (spillover)		
	Est.	*SE*	*t*	Est.	*SE*	*t*	Est.	*SE*	*t*
Antecedent Type	–0.001	0.007	–0.229	–0.018	0.008	–2.214*	0.011	0.009	1.428
Gender Match	–0.003	0.007	–0.426	–0.043	0.012	–3.551**	–0.022	0.008	–2.808**
Type × Match	–0.014	0.007	–2.127*	–0.009	0.008	–1.083	–0.001	0.008	–0.102

. *p* < 0.10; * *p* < 0.05; ** *p* < 0.01; *** *p* < 0.001.

Region 7: The analysis revealed a significant interaction of antecedent type and gender match. Planned comparisons found a marginally significant simple effect of gender match in the DP condition (Est. = –0.033, SE = 0.019, t = –1.807, p = 0.071), but none in the QP condition (Est. = 0.022, SE = 0.019, t = 1.201, p = 0.230).

Region 8: The analysis revealed significant main effects of antecedent type and gender match. Planned comparisons found a significant simple effect of gender match in both the DP condition (Est. = –0.102, SE = 0.029, t = –3.553, p < 0.001) and QP condition (Est. = –0.068, SE = 0.029, t = –2.342, p < 0.05).

Region 9: The analysis revealed a significant main effect of gender match. Planned comparisons found a significant simple effect of gender match in the DP condition (Est. = –0.047, SE = 0.023, t = –2.058, p < 0.05) and a marginally significant simple effect of gender match in the QP condition (Est. = –0.043, SE = 0.023, t = –1.919, p = 0.055).

The impact of including previous region reading times as a predictor is to shift the effects in the original analysis, making the key outcome pattern somewhat more subtle: There still is an interaction in region 7, but the GMME in the DP conditions is now only marginally significant. Additionally, the interaction disappears in region 8, where we now also find a fully significant GMME for the QP conditions. (But note that our marginally significant simple effect in the QP condition in the first analysis here already contrasts with Moulton and Han's (2018) findings, as they found no GMME for QP conditions.) Nonetheless, at a minimum, it still holds that there is a difference between QP and DP conditions, given the interaction in region 7. We can't know whether the original data would exhibit these same shifts if previous region reading times were included. For the purposes of the discussion to follow, we'll take the new analyses including the additional predictor as our main focus, though we'll summarize outcomes for analyses without that predictor in footnotes when reporting subsequent data analyses.

2.5 Discussion

Experiment 1 was conducted for the purpose of replicating Moulton and Han (2018), to ensure a sound baseline for the variations in Experiment 2. Overall, the data do generally replicate the effects in Moulton and Han, though in a more subtle way once we include the additional predictor of previous region reading times in the model. Even so, the interaction in region 7 indicates a difference in the impact of the gender manipulation based on antecedent type. This is further confirmed by the marginally significant simple effect of gender in the DP condition and the absence thereof in the QP condition. However, note that in our data, the interaction disappears in spillover regions 8 and 9, and the gender manipulation has a significant or

marginally significant simple effect there for both DP and QP conditions. This contrasts with Moulton and Han,[7] but is not, in principle, incompatible with their overall generalizations. It still reflects an initial phase where the referential condition, with no c-command requirement, involves immediate establishment of the dependency on the antecedent in a way that is sensitive to gender features, whereas effects of gender don't arise until later in the QP condition with a covarying pronoun. The difference is that we do find an effect in the reading times downstream, but since, conceptually, the gender mismatch has to matter at some point down the line (given the final interpretation of the sentence), whether or not that has repercussions in a given processing measure at some later point is largely an orthogonal question. Note that while the most extreme construal of Moulton's (2017) proposal might predict there to be no GMMEs at all, since there is no direct link between the pronoun and the candidate antecedent, one could also imagine other ways compatible with the present finding in which infelicity causing a processing delay could come into play later on in this proposal. Since the varied stimuli in Experiment 2 yield immediate GMMEs for the QP condition, we won't dwell on this issue here.

While the present results are compatible with Moulton and Han's (2018) preferred interpretation, these results – just like Moulton and Han's – are subject to a potential alternative interpretation, as they discuss: While it could be that the dependency is immediately established in the QP condition, with any potential impact of gender mismatch not unfolding until later, it also could be that the dependency itself is not established until later. If that were the case, another effect might be expected in both the Match and Mismatch conditions, namely, a (temporary) unheralded pronoun effect (Greene et al., 1994), due to encountering a pronoun without an explicit antecedent within the sentence. This would predict initial slow-downs in the Match condition for QPs relative to DPs. Moulton and Han discuss and discard this possibility in the context of their Experiment 2. There, a c-command condition, rather than replacing the QP with a DP, serves as a control, and they do not find any difference in processing time upon encountering the pronoun. However, the comparison here is less than ideal, due to structural differences between NoCC and CC sentences – the structurally parallel DP and QP sentences make for a more telling comparison. And indeed, additional post-hoc analyses, again applying the emmeans package to the model with a predictor for previous region reading times, show a significant simple effect of antecedent type in the Match condition in region 8 (Est. = –0.054, $SE = 0.023$, $t = –2.339$, $p < 0.05$), due to slower reading times in the QP Match condition.[8]

---

[7] Although we should note that Moulton and Han's (2018) experiment also contains a slight hint of a parallel effect, with a marginally significant effect ($p = 0.09$) of increased difficulty in the QP mismatch condition compared to the QP match condition in that same region.

[8] A parallel trend seems to be present, at least numerically, in the graphs for the original version of this experiment in Moulton and Han (2018).

This may reflect an unheralded pronoun effect, and therefore be evidence against the proposal that a dependency is immediately established. This could be because wide scope has not been robustly computed in early processing for non-c-commanding QP antecedents in these materials, as suggested by Kush and Eik (2019). The results of Experiment 2 will shed more light on this issue, so we return to it in 3.6 (also see Section 5). For the moment, the main take-away from this initial replication is that we can detect differences between the DP and QP conditions in terms of the emergence of GMMEs. This sets the stage for our second experiment, where we manipulate the original stimuli along the lines suggested by Kush and Eik.

3. Experiment 2

In light of their finding of immediate GMMEs in donkey sentences, where the explanation of the Moulton and Han (2018) data detailed in Moulton (2017) predicts the absence of GMMEs parallel to the NoCC condition in their work, Kush and Eik (2019) argue for a uniform processing mechanism that retrieves the antecedent in the presence of c-command or other contextual factors facilitating an anaphoric dependency. But while Kush and Eik's results clearly establish that some non-c-commanding antecedents for covarying pronouns can give rise to immediate GMMEs, it's far from clear whether this will generalize to other cases. Recall, among other things, that donkey sentences do not actually fall under Barker's (2012) scope constraint proposal, since the indefinite antecedents in donkey sentences do not take scope over the relevant pronoun sites. Thus, it remains a genuinely open question whether immediate GMMEs can arise in sentences such as those in the NoCC condition, and, in particular, whether this can be brought about by manipulating the sorts of factors suggested to be at play by Kush and Eik.

3.1 Adjustments to Experiment 1

Kush and Eik (2019, p. 12) speculate that a key factor in how quickly anaphoric dependencies are established in the relevant sentences is "how easy it is to adopt a *quantificational, distributive,* or *multi-event* reading of the fronted adjunct," and Moulton and Han's (2018) stimuli arguably do not facilitate such readings. The pragmatics of producing such a reading may be informed by the literature on telescoping. Telescoping, as mentioned earlier, is usually described as semantic binding across sentence boundaries. Certain factors have been identified as promoting felicitous telescoping (Anderssen, 2011; Poesio & Zucchi, 1992). Among these is a *scripted, non-accidental,* and/or *generic* relationship between sentences, such that there is an appearance of regular relatedness – perhaps causation or expected succession – between the events of the first sentence and the second sentence. For example, (11b) demonstrates the change in felicity of a telescoping interpretation of (11a) when a context is provided to promote a scripted reading (examples from Poesio and Zucchi).

(11) a. # Every$_i$ dog came in. It$_i$ lay down under the table.

  b. I went to the circus last night. They had a number involving dogs that went like this: The circus performers put a table on some supports. Then, every dog came in. It lay down under the table, stood on its back paws, and lifted the table with its front paws.

As these types of cues seem to promote covariation for telescoping examples, Kush and Eik (2019) propose that they will do so as well for the stimuli in Moulton and Han (2018). They propose four specific adjustments in this vein, in particular:

(12) a. Change from past tense to present tense.

  b. Add an indefinite DP to the adjunct clause.

  c. Remove the intervening raising predicate *it seems/it appears.*

  d. Appear generally scripted in nature.

For Experiment 2, we implemented adjustments along these lines for the Moulton and Han (2018) stimuli used in Experiment 1. While (12a–c) involved fairly straightforward alterations, (12d) was more open-ended, and involved modifying the content of each clause such that the events in the second clause were more closely related to those in the first. The outcome was a set of stimuli that were completely parallel to the stimuli in Experiment 1 in terms of the overall syntactic configuration, but whose salient interpretation had the features above to promote a scripted reading. (13) illustrates the resulting variants of the original stimuli in (10).

(13) a. After $_1/_2$ each boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ he $_6/_7$ goes $_7/_8$ to clean the $_8/_9$ barn and stables. *(QP Match)*

  b. After $_1/_2$ each boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ she $_6/_7$ goes $_7/_8$ to clean the $_8/_9$ barn and stables. *(QP Mismatch)*

  c. After $_1/_2$ the boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ he $_6/_7$ goes $_7/_8$ to clean the $_8/_9$ barn and stables. *(DP Match)*

  d. After $_1/_2$ the boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ she $_6/_7$ goes $_7/_8$ to clean the $_8/_9$ barn and stables. *(DP Mismatch)*

Kush and Eik (2019) argue that the changes outlined in (12) and implemented in (13) intuitively make a scripted interpretation of these sentences more easily available. The change to present tense is of particular importance, we argue (and elaborate on in 5.3), as the sentences are now more plausibly interpreted as being part of a general script of what happens whenever a boy fetches a bucket of water: That same boy goes to clean the barn and stables. This is in contrast to the past tense sentences in the original stimuli in (10), which seem to be easily construed as a simple recounting of events that have occurred, without any scripted notions. We find that the raising predicate *it seems/it appears* also promotes a circumstantial, non-scripted reading, motivating its removal. Finally, by thematically connecting the events in the subordinate and

matrix clauses, and by using an indefinite which can quantify over the events in the matrix clause, we reinforce the availability of the scripted interpretation and thus, in theory, the covarying interpretation.

Experiment 2 keeps constant the syntactic structure of Experiment 1, while contextually facilitating a scripted interpretation. If the absence or delay of a GMME is due to the general syntactic structure of the sentences and the underlying semantic mechanisms of covariation (via quantification over situations), as on Moulton and Han's (2018) account, then this variation should have no effect. In contrast, on Kush and Eik's (2019) proposal, where contextual factors matter for how likely a relevant scopal interpretation and corresponding anaphoric dependency is, we should see comparable GMMEs in both antecedent type conditions and no interaction, assuming our manipulation is sufficient and successful.

3.2 Materials and procedure

The procedure remained the same as in Experiment 1, using the adjusted stimuli. The materials were directly adapted from those in Experiment 1, with twenty test items in four conditions, as exemplified in (13). The same twenty fillers from Experiment 1 were used, with one region removed (typically a modifying adjunct) to match the number of regions in the adjusted stimuli. A link to the experiment can be found in (21) in Appendix A.

3.3 Participants

Sixty-seven undergraduates self-identifying as native English speakers were recruited through the University of Pennsylvania's subject pool, none of whom had participated in Experiment 1.

3.4 Analysis

The same data removal criteria as for Experiment 1 were applied. This eliminated three participants, leaving data from a total of 64 participants. Forty-seven individual experimental trials (4%) were removed, following the same removal criteria as in Experiment 1.

RT data were natural log-transformed and analyzed with a linear mixed-effects model, following the same approach as for Experiment 1, using the maximal random effects structures that would converge. For all three regions, this was a model with a random slope for antecedent type by participants.
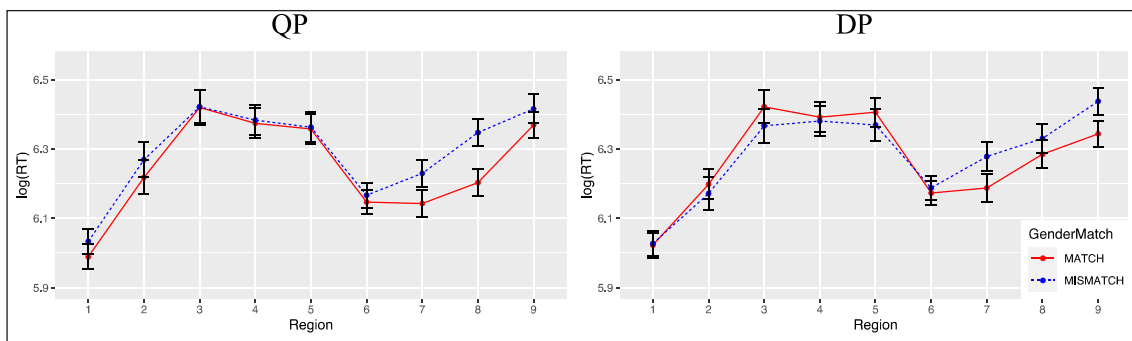
3.5 Results

Across all conditions and fillers, participants answered the comprehension questions with a mean accuracy of about 0.89 ($SE = .008$). **Table 4** shows the mean proportion of correct responses in each condition. There appears to be no major effect on comprehension from the manipulations.

Table 4: Mean accuracy rates of comprehension question responses (*SE*) in Experiment 2.

	Match	**Mismatch**
QP	.89 (.018)	.88 (.024)
DP	.90 (.017)	.88 (.022)

Figure 2 provides a graph of natural log-transformed mean RTs by region. In this experiment, region 6 is the critical one, but since it only included the pronoun in this version, the spillover regions 7 and 8 are especially important to consider.

Figure 2: Log-transformed mean RTs (with standard errors) by region in Experiment 2.

The results of the analysis are summarized in **Table 5**. We focused on models that included previous region reading times as a predictor.[9] The effects of previous region reading times were highly significant throughout, but are not reported here in detail.

Table 5: Summary of statistical analysis for Experiment 2.

	Region 6 (pronoun)			**Region 7 (spillover)**			**Region 8 (spillover)**		
	Est.	*SE*	*t*	Est.	*SE*	*t*	Est.	*SE*	*t*
Antecedent Type	0.007	0.009	0.769	0.017	0.009	1.838.	0.005	0.010	0.472
Gender Match	−0.011	0.009	−1.328	−0.041	0.009	−4.473***	−0.035	0.010	−3.523***
Type × Match	−0.003	0.009	−0.309	−0.004	0.009	−0.419	0.022	0.010	2.235*

. $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

---

[9] The results without such a predictor were essentially identical in terms of significance patterns; the only divergences were minor and not crucial to our interpretation of the data, namely: (i) in region 7, the main effect of antecedent type was significant ($p < 0.05$), rather than merely marginally significant, (ii) in region 8, the simple effect indicative of a GMME in the DP condition was still significant ($p < 0.05$), (iii) the simple effect of antecedent type in the match condition was significant ($p < 0.05$), rather than merely marginally significant.

Region 6: There are no significant effects.

Region 7: The analysis revealed a marginally significant main effect of antecedent type and a significant main effect of gender match. Planned comparisons found significant simple effects of gender match in both the QP condition (Est. $= -0.074$, $SE = 0.026$, $t = -2.857$, $p < 0.01$) and the DP condition (Est. $= -0.089$, $SE = 0.026$, $t = -3.464$, $p < 0.001$).

Region 8: The analysis revealed a significant main effect of gender match and a significant interaction of antecedent type and gender. Planned comparisons found a significant simple effect of gender match in the QP condition (Est. $= -0.114$, $SE = 0.028$, $t = -4.067$, $p < 0.001$), but not in the DP condition (Est. $= -0.026$, $SE = 0.028$, $t = -0.924$, $p = 0.356$).

3.6 Discussion

The adjustments implemented on the stimuli from Experiment 1 led to immediate GMMEs in both the QP and DP conditions, and no interaction of antecedent type and gender match of the sort found before (see below on the interaction in region 8). These results suggest that the QP antecedents, including their gender information, were as immediately and robustly accessed as the DP antecedents. The changes aiming to create a more script-like reading for a parallel syntactic structure thus did induce an interpretation where an anaphoric dependency of the pronoun on the QP antecedent was established without any delay, as predicted by Kush and Eik (2019).

Unlike in Experiment 1, but parallel to Moulton and Han's (2018) Experiment 3, GMMEs did not emerge until the first spillover region. Note, however, that unlike there, the critical region in the present study only contained the pronoun, and thus was very short, making spillover effects more likely. Moreover, prior findings by Cunnings et al. (2015), who found GMMEs beginning in the pronoun region in a DP condition in their eye-tracking data, suggest that the slightly delayed effect in the spillover region here is due to the nature of the self-paced reading paradigm (though also note that we do find GMMEs in the region containing only the pronoun in Experiment 3).

Another aspect of the data worth commenting on is the interaction between antecedent type and gender match in region 8, with a greater effect of gender mismatch in the QP condition. Note that this seems to mainly be driven by slower reading times in the DP Match condition: Post-hoc analysis of region 8 shows a marginally significant simple effect of antecedent type in the Match condition (Est. $= 0.053$, $SE = 0.028$, $t = 1.897$, $p = 0.059$). This condition may be slightly less compatible with the scripted nature of the new stimuli. Further aspects of this tentative finding – if fully substantiated statistically – need to be explored in future work. But the direction of this effect in the second spillover region goes directly counter to Moulton and Han's (2018) predictions of an absence of GMMEs for QP-anteced pronouns in the NoCC condition; i.e., if anything, the GMME for the QP conditions seems to be more pronounced and long-lasting than in the DP conditions.

Returning to the finding that is key to the main question we are pursuing, Experiment 2 crucially establishes that syntactic structure alone cannot be blamed for the delayed GMMEs for the stimuli in Experiment 1 (or for their absence in Moulton and Han's (2018) data). Experiment 2 employed exactly the same syntactic configuration, with a QP in a temporal adjunct clause, and yet the modifications of the stimuli aiming for a more scripted and non-accidental relationship between the two clauses result in immediate GMMEs. Taking the standard stance that GMMEs reflect the establishment of an anaphoric dependency, these data thus show that non-c-commanding QPs are immediately considered as antecedents here, and their gender is evaluated right when the relevant pronoun is encountered, as fast as in the DP condition, which doesn't require c-command.

There is another point worth noting in relating the findings from Experiment 2 to Moulton and Han's (2018) account and their interpretation of their data. Their proposal explicitly argued that in NoCC configurations with QPs, the dependency between the situation pronoun and its binder is indeed established immediately, but since there is no gender information encoded at this level, gender mismatch effects go initially unnoticed, until some later point when the D-Type pronoun interpretation is fully considered. Correspondingly, and as noted in the discussion of Experiment 1 above, they argued against an alternative interpretation of their effect in terms of an unheralded pronoun effect. The results of Experiment 2 speak against both of these takes. First, we do find immediate GMMEs in the QP condition. This suggests full and immediate consideration of the information encoded on the pronoun, including gender features, rather than an initial phase of merely considering the anaphoric dependency of the situation pronoun. Secondly, recall that Experiment 1 showed what can plausibly be seen as an unheralded pronoun effect, in the form of a significant simple effect in the Match conditions, with faster reading times in the first spillover region for DP than QP sentences. But no such effect is present in Experiment 2, and we even find a marginally significant effect in the opposite direction in the second spillover region. This, in turn, suggests that GMMEs are indeed tied to whether the relevant potential antecedent is considered in interpreting the pronoun.

The present results align with the predictions by Kush and Eik (2019) and are, in principle, compatible with their proposal of a mechanism that uniformly governs the retrieval of QP antecedents. However, the details of just how the scripted interpretation and the overall manipulation of contextual factors feed into such a mechanism are yet to be spelled out. Furthermore, the fact remains that c-command and non-c-command configurations seem to differ in that only the latter are affected by the present contextual manipulations. We will return to these more general issues about how to interpret the data in Section 5. But first, it is worth testing for any potential remaining differences empirically. Here, we follow suit with the previous literature in also comparing c-commanding and non-c-commanding QP antecedents in as minimally varied sentences as possible, as in Experiment 2 from Moulton and Han (2018). Using the modified

stimuli from Experiment 2 (with additional c-command variations), Experiment 3 will allow us to assess more directly whether the non-c-commanding antecedents there exhibit any differences in processing time-course from c-commanding antecedents, as this is not directly ruled out yet by our findings so far.

4. Experiment 3

While the previous experiment has shown that syntactic structure alone cannot be held responsible for processing delays in general, the possibility remains that c-command does play a role in how easily QP antecedents are retrieved. By adding certain contextual pressures to exceptionally covarying sentences, QP antecedents become more readily accessible, according to Experiment 2. Are they, however, as readily accessible as *c-commanding* QP antecedents?

Experiment 3 utilizes the QP sentences (Match and Mismatch conditions) from Experiment 2, labeled NoCC in (14), and adds c-commanding (CC) condition variants. Parallel to Moulton and Han's (2018) Experiment 2, these were implemented by moving the temporal conjunction (*After/Before/When*) to the region immediately preceding the pronoun. To maintain a parallel event structure and overall interpretation between the CC and NoCC conditions, *before* was changed to *after* (and vice versa) in relevant variants. One challenge in creating stimuli varying the roles of the main clause and adjunct clause, for the two clauses in play, is keeping the quantifier and pronoun in place in terms of their surface position, which is important for maintaining a constant distance between them. Moulton and Han's stimuli achieve this by adding the embedding with *it seems*, which can appear in either clause. However, this embedding is hard to integrate with the manipulations in our Experiment 2. As an alternative solution, we add a fronted adjunct, usually an adverb or a PP, to the CC sentences. The same adjunct is also added to the NoCC condition in the position immediately preceding the pronoun. The goals guiding the particular adjunct choices were chiefly the following: (i) maintain the scripted nature of the sentence, to ensure the contextual cues remained intact; (ii) make adjunction to the verb phrase in the NoCC condition plausible; (iii) limit the number of syllables of the adjunct to three, to match pronoun-antecedent distance between conditions as closely as possible. To meet (ii), in particular, and prevent potential garden-pathing, the adjunct was followed by a comma. The resulting sentences are exemplified in (14).[10]

---

[10] It is possible that introducing the additional adjunct, which can make the sentences somewhat cumbersome in certain cases, introduces a confound of its own. Appendix B summarizes a version of this experiment, not reported here, that utilizes the same stimuli with the adjunct removed (and antecedent-pronoun distance correspondingly varying slightly). The results are comparable to those in Experiment 3, suggesting that neither the antecedent distance nor the addition of the adjunct crucially contributes to the relevant aspects of the results. Note that in this variant, the NoCC did not have a comma preceding the pronoun, matching the CC condition in this regard, suggesting that the comma in the NoCC sentences in (14) did not crucially affect the results.

(14) a. After $_1/_2$ each boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ on foot, $_6/_7$ he $_7/_8$ goes $_8/_9$ to clean the $_9/_{10}$ barn and stables. *(NoCC Match)*

  b. After $_1/_2$ each boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ on foot, $_6/_7$ she $_7/_8$ goes $_8/_9$ to clean the $_9/_{10}$ barn and stables. *(NoCC Mismatch)*

  c. On foot, $_1/_2$ each boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ before $_6/_7$ he $_7/_8$ goes $_8/_9$ to clean the $_9/_{10}$ barn and stables. *(CC Match)*

  d. On foot, $_1/_2$ each boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ before $_6/_7$ she $_7/_8$ goes $_8/_9$ to clean the $_9/_{10}$ barn and stables. *(CC Mismatch)*

If c-command has an independent role to play in the accessibility of QP antecedents in covarying constructions, we should expect to see the NoCC conditions exhibit reliably greater retrieval costs, such that there is an interaction between the structure type and gender match conditions, at a minimum in early regions. Alternatively, if c-command has no privileged role in processing in the present sentence variants, we should see no such interaction, but rather consistent early GMMEs across conditions, parallel to Experiment 2.

4.1 Materials and procedure

The procedure followed that of the previous experiments. There again were twenty test items in four conditions, as exemplified in (14). Each condition had either a NoCC structure or a CC structure, and either a gender match or gender mismatch between the pronoun and its possible antecedent. The same twenty fillers from Experiment 1 were used. A link to the experiment can be found in (22) in Appendix A.

4.2 Participants

Seventy-five undergraduates self-identifying as native English speakers took part in the experiment, none of whom had participated in Experiments 1 or 2 before.

4.3 Analysis

Following the same data removal criteria as for the previous experiments, five participants' data were removed, leaving a total of 70 participants for analysis. Forty-two individual experimental trials (3%) were removed, following the same removal criteria as in the previous experiments.

Analysis was carried out on natural log-transformed RT data, using a linear mixed-effects model with the maximal random effects structure that would converge, as before. The model for region 7 had by-participant random slopes for gender match and structure, and an uncorrelated by-item slope for gender match; the one for region 8 had the same by-participant random slopes

and a random intercept only for items; the only random slope for region 9 was for gender match by participants.

4.4 Results

Across all conditions and fillers, participants answered the comprehension questions with a mean accuracy of 0.92 (*SE* = 0.007). **Table 6** shows the mean proportion of correct responses by condition, with no major effect of condition on comprehension.

Table 6: Mean accuracy rates of comprehension question responses (*SE*) in Experiment 3.

	Match	**Mismatch**
NoCC	.91 (.019)	.89 (.017)
CC	.92 (.014)	.89 (.019)

Figure 3 provides a graph of natural log-transformed mean RTs for each region. The critical region is region 7 (again, just containing the pronoun), and regions 8 and 9 are possible spillover regions.
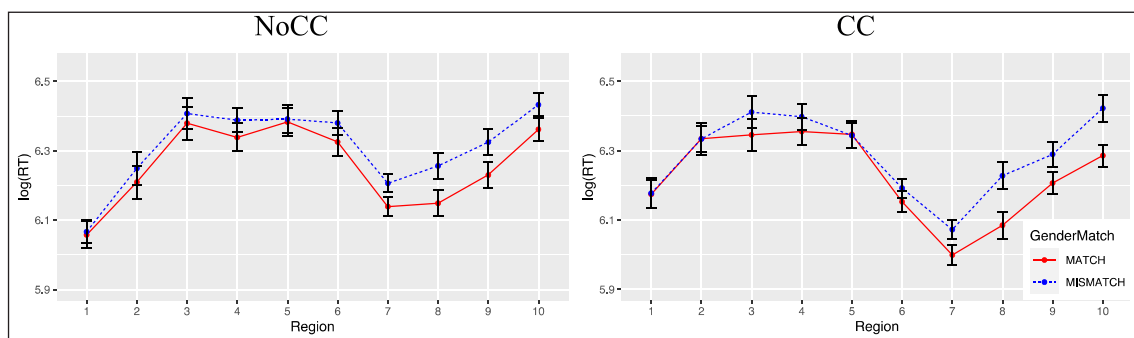

Figure 3: Log-transformed mean RTs (with standard errors) by region in Experiment 3.

The results of the analysis are summarized in **Table 7**. We focused on models that included previous region reading times as a predictor.[11] The effects of previous region reading times were highly significant throughout, but are not reported here in detail.

---

[11] The results without such a predictor were essentially identical in terms of significance patterns; the only divergences were minor and not crucial to our interpretation of the data, namely: (i) in region 8, there was a marginally significant main effect of structure ($p = 0.057$), as opposed to no effect; (ii) in region 9, there still was a significant simple effect indicating a GMME in the CC conditions ($p < 0.001$), as opposed to no effect.

Table 7: Summary of statistical analysis for Experiment 3.

	Region 7 (pronoun)			Region 8 (spillover)			Region 9 (spillover)		
	Est.	*SE*	*t*	Est.	*SE*	*t*	Est.	*SE*	*t*
Structure Type	–0.044	0.008	–5.622***	0.012	0.011	1.103	–0.007	0.008	–0.829
Gender Match	–0.028	0.008	–3.435**	–0.047	0.009	–5.124***	–0.023	0.008	–2.696**
Type × Match	–0.002	0.007	–0.282	–0.014	0.008	–1.781.	0.005	0.008	0.550

. $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Region 7: The analysis revealed significant main effects of structure type and gender match. Planned comparisons found significant simple effects of gender match in both the NoCC condition (Est. = –0.053, *SE* = 0.021, *t* = –2.467, $p < 0.05$) and the CC condition (Est. = –0.060, *SE* = 0.022, *t* = –2.811, $p < 0.01$).

Region 8: The analysis revealed a significant main effect of gender match and a marginally significant interaction of structure type and gender match. Planned comparisons found significant simple effects of gender match in both the NoCC condition (Est. = –0.065, *SE* = 0.024, *t* = –2.709, $p < 0.01$) and the CC condition (Est. = –0.122, *SE* = 0.024, *t* = –5.015, $p < 0.001$).

Region 9: The analysis revealed a significant main effect of gender match. Planned comparisons found a significant simple effect of gender match in the NoCC condition (Est. = –0.055, *SE* = 0.023, *t* = –2.329, $p < 0.05$), but none in the CC condition (Est. = –0.037, *SE* = 0.024, *t* = –1.531, $p = 0.127$).

4.5 Discussion

We find GMMEs in both the NoCC and CC conditions in the critical and first spillover region, and also in the second spillover region for NoCC. In contrast to Moulton and Han's (2018) Experiment 2, with syntactic structures overall parallel to ours here, our modified stimuli, which aim to set up a more scripted relation between events, do not give rise to an interaction of structure type and gender match in any of the regions. As such, these results show that QP antecedents in exceptionally covarying constructions with a QP in a temporal adjunct clause can be as easily and as quickly processed as QP antecedents that c-command a pronoun, once the overall interpretation is more supportive of a covarying interpretation.

Parallel to what we saw in Experiment 2, we find early significant GMMEs in the NoCC condition. Here, these even start in the critical region containing just the pronoun, in contrast

to Experiment 1, where GMMEs only emerged in the spillover region across conditions. There is no indication of a delay in GMMEs relative to controls (here, the CC condition; DP antecedents in Experiment 2). Thus, the presence or absence of c-command of a pronoun by a potential QP antecedent does not, in general, determine the time-course of GMMEs. Both configurations can exhibit immediate effects, although in the NoCC condition, this is further modulated by contextual and general interpretive properties of the sentences.

Before discussing the broader theoretical repercussions of the present set of findings and comparing them to previous work, a caveat is in order. The absence of initial interactions here is, of course, not fully conclusive, in that it is limited to the still relatively course-grained method of SPR at hand. There could be smaller-scale timing discrepancies between the CC and NoCC conditions that are too fine-grained to be captured here. Some potential hints of this can be found in the data, e.g., in region 8, where a marginally significant interaction appears to be due to a numerical difference in the size of the GMMEs. These are slightly larger in the CC condition, with 80 ms, than in the NoCC condition, with 69 ms. At the same time, the fact that we already find GMMEs in the region containing only the pronoun in Experiment 3 (an effect that replicates in the variant of the experiment reported in Appendix B) points to rather rapid effects. Furthermore, in some cases, we find more extended or pronounced GMMEs for the QP and NoCC conditions. On balance, and in the absence of positive evidence for differences due to the presence or absence of c-command, we will proceed to our general discussion, taking the present result patterns at face value.

5. General discussion

Our experiments use variations of the stimuli from Moulton and Han (2018) with QP antecedents in temporal adjunct clauses. They implement suggestions by Kush and Eik (2019) aimed at making covarying interpretations of pronouns in these sentences more easily accessible. Our main empirical question was whether these pronouns would then give rise to GMMEs. In theoretical terms, we are interested in how the empirical findings inform theories of pronoun processing and antecedent retrieval. At the same time, we're also considering implications for the grammatical mechanisms that give rise to covarying interpretations of pronouns in different configurations. In this section, we sum up our findings and discuss them in light of the broader questions at hand.

5.1 Can non-c-commanding QP antecedents be accessed early in processing?

In the present context, this question boils down to whether we find GMMEs for non-c-commanding QP antecedents, relative to controls such as DP antecedents or c-commanding QP antecedents, or whether these are absent or delayed. Moulton and Han's (2018) original

findings, as well as our replication in Experiment 1, suggest the latter. Moulton and Han found no GMMEs in such cases. Our initial experiment replicates this, with an initial interaction due to GMMEs for DP antecedents, but not QP antecedents. However, we do find an effect of gender in the spillover regions for QPs that suggests a delay, rather than a complete absence of GMMEs. Nonetheless, the pattern in Experiment 1, on its own, is compatible with the generalization Moulton and Han build on: There is at least an initial phase where GMMEs for pronouns with non-c-commanding antecedents fail to arise.

The data from Kush and Eik (2019) on GMMEs in donkey sentences provided a first contrast to this, as GMMEs with non-c-commanding QP antecedents in that construction arise as quickly as for DP antecedents. Our data from Experiments 2 and 3 on sentences structurally parallel to ones from Moulton and Han's (2018) experiments extend this to temporal adjunct clauses: In Experiment 2, DP and QP antecedents give rise to GMMEs in the same regions, and there's no interaction between them. Experiment 3 further confirms that non-c-commanding QP antecedents lead to GMMEs as quickly as c-commanding ones.

The shift in GMMEs between the original and modified stimuli appears to be due to the changes based on suggestions from Kush and Eik (2019). These aimed to give the stimuli a more scripted and non-accidental interpretation. While suggesting that non-c-commanding QP antecedents *can* be accessed quickly, the overall data show that this is subject to variation based on non-structural aspects of the stimuli.

5.2 What are the implications for c-command in theories of pronoun processing?

One class of theories on processing pronouns relative to QP antecedents posits a special role for c-command. The most straightforward variant of this is that the processor's search space for antecedents is strictly restricted in purely structural terms. On that view, only noun phrases in syntactic positions that c-command the pronoun are considered, at least during an initial phase. In particular, we considered Moulton and Han's (2018) proposal along these lines, spelled out in technical detail in Moulton (2017). Here, covariation in non-c-commanding environments, analyzed as involving binding of situation variables, is claimed to not (or at least not immediately) involve evaluation of gender features. This is because the pronoun for which an antecedent is sought is a situation pronoun without such features. Our findings from Experiments 2 and 3 are inconsistent with this general type of approach, as non-c-commanding QP antecedents there do give rise to GMMEs as quickly as the relevant controls. This shows that the QP antecedents in question are accessed early on. It thus speaks against an initial processing phase where only c-commanding QP antecedents are considered, adding to the previous evidence from donkey sentences in Kush and Eik (2019). More generally, this drives

home the point that the presence or absence of GMMEs is not determined by purely structural properties of the constructions alone. Correspondingly, theories about the search space for antecedent retrieval can't be defined in purely structural terms.

To some extent, this conclusion contrasts with other findings concerning principles of the binding theory, especially for reflexives. For those, it has been argued that structurally illicit antecedents are categorically ignored in early processing (e.g., Chow et al., 2014). It was this work that inspired Kush et al. (2015) and Cunnings et al. (2015) to test for similar effects due to the c-command constraint in QP binding that Reinhart (1983) proposed. Whether or not the former findings turn out to hold in full generality, with possible limitations, e.g., in light of non-standard uses of reflexives, it is clear from the present results that the range of potential noun phrase antecedents for bound pronouns is not, in general, limited to ones that c-command the relevant pronoun, not even during an initial processing phase.[12] Further confirmation of this finding should be sought in future work on the processing of stimuli like those in Experiments 2 and 3, e.g., using eye-tracking to obtain data with a higher temporal resolution than self-paced reading. Alternative research paradigms (e.g., see Badecker & Straub, 2002) may also shed more light on the way in which c-command and contextual pressures interact in retrieval, by measuring the time-course of processing multiple candidate antecedents.

Our data clearly show that non-c-commanding QP antecedents *can* be accessed quickly. But there still remains a difference between non-c-commanding QP antecedents and the c-commanding ones: For the former, GMMEs only arose right away once adjustments to the original stimuli were made. The latter showed the effects regardless of the version of the stimuli. This variation in the presence of the effect, then, leaves open the question of whether c-command has a special role in the processing of pronouns with QP antecedents after all, and if so, what that role might be. Considering what might modulate GMMEs in non-c-commanding cases (see 5.3 for more detail), such a role is likely to be a very indirect one at best. The most plausible possibility is that whether or not a QP antecedent is accessed at a given point in time depends on whether or not it is interpreted as having semantic scope over the pronoun.[13] If this is on the right track, then what's special about c-commanding QP antecedents might simply be that they are interpreted as taking scope over expressions in their c-command domain by default: C-command does generally align with surface scope interpretations, and these, in turn, have been argued to be preferred in processing (Anderson, 2004). Thus, while semantic scope does

---

[12] Another early finding along parallel lines comes from research on VP-ellipsis, which argued that binding in that construction can occur in the absence of c-command when certain intonational signals are present (Hirschberg & Ward, 1991).

[13] Note that we're considering this here only from the perspective of what constraints might guide the processor in accessing antecedents. This is a separate question from whether or not a grammatical account in terms of scope, without any role for c-command, for all covarying pronouns is theoretically warranted. See 5.4 for more details.

not imply c-command, as in the temporal adjunct clauses in our experiments, c-command does quite generally imply the easy availability of corresponding semantic scope. The early GMMEs in Moulton and Han's (2018) original c-command conditions and our Experiment 1, which do not promote more scripted interpretations, would then be explained in terms of the relevant scope being available by default. This, in turn, would be based on independent processing principles, making the relevant interpretation independent of contextual support. In contrast, the QPs in temporal adjunct clauses in the no-c-command conditions do seem to require additional support to make such a scope interpretation easily and quickly available. The broader prediction that follows from this is that the presence of GMMEs should correlate with the availability of relevant scopal interpretations, as independently measured. Testing this more generally seems like a formidable task to be taken up in future work.

The present study implemented the changes to Moulton and Han's (2018) materials suggested by Kush and Eik (2019), and our findings directly align with the latter authors' prediction for them. The overall empirical picture is, therefore, entirely consistent with the uniform antecedent retrieval mechanism these authors propose. They extend a cue-based framework to characterize the nature of this mechanism. Cue-based models traditionally assume that the search for an antecedent is guided by item-specific features intrinsic to the items being retrieved, like morphological features (Lewis et al., 2006). Each word encountered by the parser triggers specific retrieval cues, which guide the parser to rapidly form dependencies with items matching in appropriate features (e.g., gender or number; McElree, 2000). Kush et al. (2015), building on Kush (2013), propose to capture relational constraints in a cue-based framework by positing an ACCESSIBLE feature whose value can be dynamically updated as the parse unfolds. In the case of a c-command constraint, a given candidate QP antecedent changes status from being accessible to being inaccessible when the parse reaches a stage that is outside of the QP's c-command domain. That QP then is no longer considered as an antecedent for subsequently encountered pronouns. In Kush and Eik's (2019) uniform antecedent retrieval mechanism, the setting of the ACCESSIBLE feature can also be affected by contextual considerations relevant to determining the scope of potential antecedent QPs. Our finding of GMMEs with non-c-commanding QPs adds to their results on donkey sentences, further informing what factors impact the setting of this ACCESSIBLE feature. Since the former, but not the latter, are a case of the QP taking scope over the position of the pronoun, our results broaden our understanding of antecedent retrieval in processing across different cases of exceptional covariation. To the extent that GMMEs in donkey sentences are not subject to the same type of contextual variation that we found in temporal adjunct clauses, this further aligns with the notion that the availability of the relevant scopal interpretation in the latter is associated with the presence of GMMEs there. Future research should extend this approach to further types of exceptionally covarying constructions, such as the various other constructions documented in Barker (2012).

5.3 What modulates the availability of antecedents and the presence of GMMEs?

While the present data are indeed compatible with a uniform antecedent retrieval mechanism, we hasten to note that the details of how this retrieval mechanism works have yet to be spelled out. In particular, the way in which a host of different factors of quite different nature affect the setting of the ACCESSIBLE feature remains to be explored. The key questions in the present context are: (i) Why is there variation in the accessibility of QP antecedents in the temporal adjunct configuration, but not in c-command configurations? And (ii) what, exactly, is the nature of the variation in the former? As already noted briefly above, it is plausible that, in general, variation in QP antecedent accessibility is modulated by the relative availability of an interpretation where the QP takes scope over the pronoun. With regards to (i), the source of the variation does not seem to lie in differences in global plausibility of the relevant scopal interpretation. After all, the CC and NoCC sentence variants in Moulton and Han's (2018) study (illustrated in (7) above) essentially convey the same propositional meaning (given the switch between *before* and *after*). Rather, the structural configuration seems to affect how easily a given scope interpretation is available. Specifically, when the potential antecedent c-commands the pronoun, this corresponds to a surface scope interpretation. These interpretations, in turn, have been argued to be more easily available in general (Anderson, 2004). In contrast, in non-c-commanding configurations, such as the temporal adjunct clauses in our stimuli, the relevant scope may well be available in principle, but does not, in general, constitute the default choice. Instead, its availability is modulated by other factors.

This brings us to (ii), i.e., the question of exactly how our modification of the stimuli made an interpretation where the QP takes scope over the pronoun more easily accessible. While we aren't in a position to present a fully fleshed out formal analysis, we offer some more detailed speculations about a plausible-seeming account here. In particular we suggest linking this to the semantics involved in the change from past to present tense. Kush and Eik's (2019) intuition was that covarying readings are more easily available for quantificational, multi-event readings of these types of sentences. We suggest that this, at least in large part, results from the impact of the quantificational force of the tense operator on the relative availability of an interpretation where the quantifier in the temporal adjunct takes scope over the position of the pronoun. Consider the sketch of an analysis in (15b) of the meaning of (15a), from Moulton (2017), which roughly follows the semantic analysis of temporal adjunct clauses in Artstein (2005):[14]

---

[14] Note that modeling tense in terms of quantification over situations here makes it possible to consider a standard D-Type analysis of the pronoun, which is interpreted as *the boy* relative to the situations temporally specified by tense.

(15) a. After each boy came home, he practiced piano.
     b. ∀x[boy(x) → ∃s[came.home(x)(s) & ∃s'[practiced(hes')(s') & after(s)(s')
       & Match(x)(s')]]]

While omitting various details, this captures the following episodic meaning: For each boy *x*, there exists some situation *s* in which *x* came home, and there is a matching subsequent situation *s'* in which the relevant boy practices piano.[15] On a situation-based D-Type analysis, *hes'* here stands for a covert definite description, effectively, *the boy in s'*. This allows for a covarying interpretation without the pronoun being directly bound by *each boy*, as on a two-grammatical-mechanism view. The resulting reading is a generalization about boys, such that for each of them, a certain sequence of events is said to have occurred once. What changes when we switch to present tense is that the relevant temporal quantification becomes universal as well, which changes the logical configuration:[16]

(16) a. After each boy comes home, he practices piano.
     b. ∀x ∀s [[boy(x) & comes.home(x)(s)] → ∃s'[practice(hes')(s') & after(s)(s')
       & Match(x)(s')]]

With the two universals – quantifying over boys and situations – taking highest scope together, this now becomes a generalization over what happens when boys come home: All boy-home-coming situations are said to be followed by a situation of the relevant boy practicing piano. We think it's plausible that such a generalization is more natural and cognitively more easily accessible, and that this plays a crucial role in facilitating fast access to a covarying interpretation in our stimuli in Experiment 2. Note that any difference in accessibility across these variants has to be seen in relation to alternative scopings, which we assume the grammar makes equivalently available for both configurations. To illustrate, consider the following variations without pronouns, but with an indefinite, to explore the different scope interpretations more directly:

---

[15] *Match* here represents a matching function in the sense of Rothstein (1995), which she argues to be required even for simple temporally quantified sentences such as (i):

(i) Each time the door bell rang, Sue opened the door.

This sentence conveys that Sue opened the door as many times as the door bell rang. But if one modeled its truth conditions simply as *for all door bell ringings x, there exists a door opening by Sue y*, this would be too weak, in that the truth conditions would be satisfied by a situation where, say, after a total of 15 door bell ringings, Sue opened the door one time. The matching function *Match* ensures that there is a different door opening for each time the door bell rang.

[16] At least if we simplify the generic, habitual interpretation of the English present tense, which seems to allow for exceptions, and thus is not fully universal, but we leave it at that for present purposes. As noted by an anonymous reviewer, it may be less clear how, exactly, this facilitates the relevant scope, if we adopt a more fleshed out analysis of the present tense here; since our proposal here is relatively speculative to begin with, we leave further exploration of this issue for future discussion.

(17) a. After each boy came home, a snack was served.

b. ∃s[∀x[boy(x) → came.home(x)(s)] & ∃s'∃y[snack(y) & served(y)(s')
& after(s)(s') & Match(x)(s')]]

c. ∀x[boy(x) → ∃s[came.home(x)(s) & ∃s'∃y[snack(y) & served(y)(s')
& after(s)(s') & Match(x)(s')]]]

(18) a. After each boy comes home, a snack is served.

b. ∀s [∀x [boy(x) → comes.home(x)(s)] → ∃s'∃y[snack(y)& serve(y)(s')
& after(s)(s') & Match(x)(s')]]

c. ∀x ∀s [[boy(x) & comes.home(x)(s)] → ∃s'∃y[snack(y)& serve(y)(s')
& after(s)(s') & Match(x)(s')]]

The (b) variants here, where *each boy* does not take scope over *a snack*, require there to be one snack to be served once all the boys are home. The (c) variants, in contrast, require there to be different individual snack servings upon the arrival of each single boy. The key claim based on Kush and Eik's (2019) suggestion and our proposed explanation of the effect of the change to present tense is that in (17), interpretation (c) is less prominent and accessible relative to (b), as compared to (18), where interpretation (c) is more readily available. This, naturally, is subject to more rigorous empirical assessment, but it strikes us as intuitively plausible. As to a potential explanation of this contrast, should it indeed be confirmed, we'll offer the speculation that this has to do with a preference for letting two universal quantifiers scope together. This could be because the overall semantic representation winds up simpler in some regard, or because there is an advantage in conceptual simplicity or naturalness for such a configuration. Naturally, this line of reasoning is subject to further development and investigation, but we hope to have offered a useful first step in that direction.

5.4 Does our GMME processing data differentiate grammatical theories of covariation?

The final question is what the current empirical picture means for theories of the grammatical mechanisms underlying covariation. The results and their interpretation from Moulton and Han (2018) owed their intrigue to the notion that they provide processing evidence supporting the two grammatical mechanisms of covariation view, with a privileged role for c-commanding antecedents in processing. This was in line with a contrast in the grammar between *bona fide* binding at play in c-command configurations and an alternative, situation-based mechanism at play in non-c-command configurations. But the fuller empirical picture that is now emerging provides a more nuanced and complex picture. What our new data clearly show is that there is not a general restriction in early phases of processing to only consider c-commanding antecedents (with the standard caveat about potential limitations due to the relatively course-

grained nature of SPR measures). Once a quantifier in a temporal adjunct clause is easily interpreted as scoping over the matrix clause (due to a variety of factors, though, probably most importantly, the switch to present tense), GMMEs arise as early as for c-command and referential DP control variants. Thus, access to the relevant scope interpretation does seem to play a crucial role for the availability of a covarying interpretation of the pronoun.

One might take the apparently central role of scope here to align well with Barker's (2012) proposal that a scope constraint is all that is needed in terms of grammatical restrictions on covariation. However, the absence of a general timing difference in accessing c-commanding and non-c-commanding antecedents does not, in and of itself, speak in favor of positing a single grammatical mechanism to be at play. Oversimplifying somewhat, the choice between the two theories comes down to variants of the following semantic representations, where *hes'* in (19b) again is understood to stand for *the boy in s'*:

(19)   a.   After each boy comes home, he practices piano.

       b. $\forall s \forall x[[\text{boy}(x) \ \& \ \text{comes.home}(x)(s)] \rightarrow \exists s'[\text{practice}(\mathbf{hes'})(s') \ \& \ \text{after}(s)(s') \ \& \ M(x)(s')]]$

       c. $\forall s \forall x[[\text{boy}(x) \ \& \ \text{comes.home}(x)(s)] \rightarrow \exists s'[\text{practice}(\mathbf{x})(s') \ \& \ \text{after}(s)(s') \ \& \ M(x)(s')]]$

There does not seem to be any principled reason leading us to expect that one of these candidate pronoun interpretations should be linked to a slower or fundamentally different cognitive process in comprehension.[17] But then the fact that in this version of temporal adjunct clauses, we get GMMEs as quickly as in DP antecedent and c-commanding QP controls does not speak for or against either one of the theoretical mechanisms behind covarying pronoun interpretations entertained in the literature. Thus, we see no general reason to favor either a one-mechanism or two-mechanism grammatical theory of covariation based on the reading time data considered here.

5.5 Conclusion

Perhaps the clearest lesson from the present enterprise is that the processing mechanisms of retrieving candidate quantificational antecedents that a given pronoun covaries with are guided by rather deep and subtle aspects of the sentence's interpretation as it unfolds incrementally. In particular, the semantically subtle, and structurally innocuous, variations between Moulton and Han's (2018) stimuli and our Experiments 2 and 3 seem to directly and immediately affect how accessible a potential QP antecedent is. This is arguably due to the availability of an

---

[17] Moulton and Han (2018) were, of course, in a very different position, given their data, in that they found differences and offered to ground them in subtle theoretical distinctions. The data become much harder to interpret in such a way once we're no longer dealing with a general processing pattern correlated with the relevant structural configurations.

interpretation where the QP takes scope over the pronoun. Thus, the processes involved are not merely formal and mechanical, in terms of considering specific syntactic domains and checking for formal features; rather, they engage deeply with the compositional semantic interpretation, including whatever grammatical mechanisms one favors to deal with deviations from surface scope. We therefore submit that work on GMMEs in pronoun processing and theories of antecedent retrieval more generally should embrace these intricacies of the semantics. They should then explore the full richness of the hypothesis space that emerges as we consider theoretical and processing questions in this domain in a fully integrated perspective, including the subtleties of both structure and meaning.

Appendix A: Experiment demos

(20)   Experiment 1: https://farm.pcibex.net/r/WRsDZO/

(21)   Experiment 2: https://farm.pcibex.net/r/BooSRB/

(22)   Experiment 3: https://farm.pcibex.net/r/ngjyma/

Appendix B: Experiment 3 variant

A variant of Experiment 3 was run using stimuli that lacked the additional adjunct that had been added to the Experiment 3 stimuli to maintain a constant antecedent-pronoun distance between conditions. The QP sentences in Experiment 2, shown in (13a) and (13b) above, are left unchanged in the stimuli for this variant of the experiment, labeled NoCC in (23). In these sentences, the clausal role manipulation leads to a one-region difference in antecedent-pronoun distance between the NoCC and CC conditions. The analysis nevertheless yields overall patterns parallel to those reported in Experiment 3. These are summarized below.

(23)   a.   After $_1/_2$ each boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ he $_6/_7$ goes $_7/_8$ to clean the $_8/_9$ barn and stables. *(NoCC Match)*

          b.   After $_1/_2$ each boy $_2/_3$ fetches a bucket $_3/_4$ of water $_4/_5$ from the well $_5/_6$ she $_6/_7$ goes $_7/_8$ to clean the $_8/_9$ barn and stables. *(NoCC Mismatch)*

          c.   Each boy $_1/_2$ fetches a bucket $_2/_3$ of water $_3/_4$ from the well $_4/_5$ before $_5/_6$ he $_6/_7$ goes $_7/_8$ to clean the $_8/_9$ barn and stables. *(CC Match)*

          d.   Each boy $_1/_2$ fetches a bucket $_2/_3$ of water $_3/_4$ from the well $_4/_5$ before $_5/_6$ she $_6/_7$ goes $_7/_8$ to clean the $_8/_9$ barn and stables. *(CC Mismatch)*

We followed the same data removal criteria as in previous experiments. Three participants' data were removed, leaving a total of 73 participants for analysis. Thirty-six individual experimental trials (2%) were removed, following the same removal criteria as in the previous experiments.

Analysis was carried out on natural log-transformed RT data, using a linear mixed-effects model with the maximal random effects structure that would converge, as before. Region 6 used a model with by-participant random slopes for structure and gender match, and a by-item random slope for gender match; for region 7, there was an uncorrelated by-participant random slope for structure and an uncorrelated by-item random slope for gender match; and for region 8, there were by-participant random slopes for structure and gender match as well as their interaction, and an uncorrelated by-item random slope for structure.

Across all conditions and fillers, participants answered the comprehension questions with a mean accuracy of 0.92 ($SE = 0.007$). **Table 8** shows the mean proportion of correct responses by condition, showing no major effect of condition on comprehension.

Table 8: Mean accuracy rates of comprehension question responses (*SE*).

	Match	Mismatch
NoCC	.92 (.017)	.89 (.022)
CC	.90 (.015)	.89 (.021)

Figure 4 shows a graph of natural log-transformed mean RTs for each region. The critical region is region 6, with regions 7 and 8 as possible spillover regions.
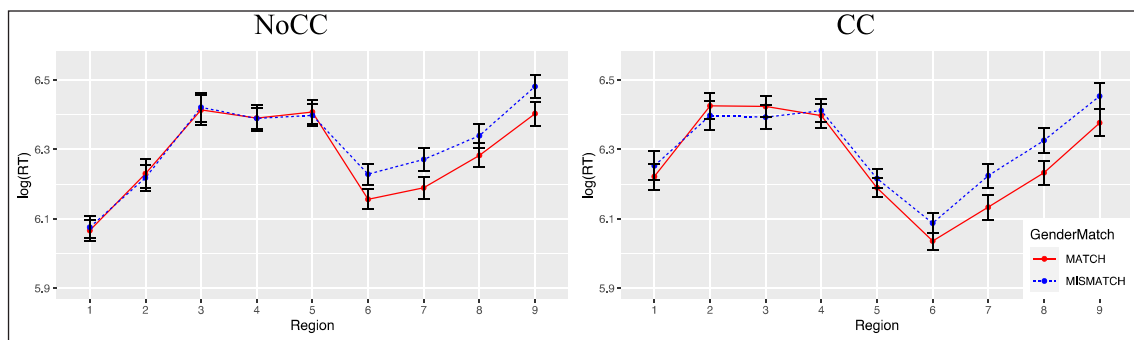
Figure 4: Log-transformed mean RTs (with standard errors) by region.

The results of the analysis are summarized in **Table 9**. We focused on models that included previous region reading times as a predictor.[18] The effects of previous region reading times were highly significant throughout, but are not reported here in detail.

Table 9: Summary of statistical analysis.

	Region 6 (pronoun)			Region 7 (spillover)			Region 8 (spillover)		
	Est.	*SE*	*t*	**Est.**	*SE*	*t*	**Est.**	*SE*	*t*
Structure Type	–0.049	0.008	–6.086***	–0.005	0.009	–0.593	–0.008	0.010	–0.790
Gender Match	–0.030	0.009	–3.355**	–0.033	0.009	–3.602***	–0.023	0.009	–2.642**
Type × Match	0.007	0.007	0.928	–0.001	0.008	–0.159	–0.007	0.009	–0.752

. $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

---

[18] The results without such a predictor were essentially identical in terms of significance patterns; the only divergences were minor and not crucial to our interpretation of the data, namely: (i) in region 7, there was a significant main effect of structure ($p < 0.01$); (ii) in region 8, there was a marginally significant main effect of structure ($p = 0.088$); (iii) in region 8, there was a marginally significant simple effect indicating a GMME in the NoCC condition ($p = 0.057$), as opposed to no effect.

Region 6: The analysis revealed significant main effects of structure type, with longer reading times in the NoCC condition, and of gender match (in the expected direction). Planned comparisons found significant simple effects of gender match in the NoCC condition (Est. = –0.073, $SE = 0.023$, $t = –3.210$, $p < 0.01$) and in the CC condition (Est. = –0.047, $SE = 0.023$, $t = –2.043$, $p < 0.05$).

Region 7: The analysis revealed a significant main effect of gender match. Planned comparisons found significant simple effects of gender match in the NoCC condition (Est. = –0.064, $SE = 0.025$, $t = –2.561$, $p < 0.05$) and in the CC condition (Est. = –0.069, $SE = 0.025$, $t = –2.764$, $p < 0.01$).

Region 8: The analysis revealed a significant main effect of gender match. Planned comparisons revealed a significant simple effect of gender match in the CC condition (Est. = –0.061, $SE = 0.026$, $t = –2.373$, $p < 0.05$), but not in the NoCC condition (Est. = –0.033, $SE = 0.025$, $t = –1.309$, $p = 0.195$).

Data accessibility statement

Stimuli lists, experiment codes, datasets, and analysis scripts can be found at https://osf.io/bvjtm/.

Ethics and consent

This study was conducted within an IRB protocol (approval number 811457) at the University of Pennsylvania, and participants provided consent for their participation.

Acknowledgements

Competing interests

The authors have no competing interests to declare.

Author contributions

Nikhil V. Lakhani: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing

Florian Schwarz: Conceptualization, Formal analysis, Methodology, Resources, Supervision, Writing – original draft, Writing – review and editing

References

Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity* (Publication No. 3156563) [Doctoral dissertation, Northwestern University]. ProQuest Dissertations & Theses Global.

Anderssen, J. (2011). *Quantification, misc.* [Doctoral dissertation, University of Massachusetts, Amherst]. ScholarWorks@UMassAmherst. https://scholarworks.umass.edu/open_access_dissertations/430/

Artstein, R. (2005). Quantificational arguments in temporal adjunct clauses. *Linguistics and Philosophy*, *28*(5), 541–597. DOI: https://doi.org/10.1007/s10988-005-6921-6

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748–769. DOI: https://doi.org/10.1037/0278-7393.28.4.748

Barker, C. (2012). Quantificational binding does not require c-command. *Linguistic Inquiry*, *43*(4), 614–633. DOI: https://doi.org/10.1162/ling_a_00108

Barker, C., & Shan, C. (2008). Donkey anaphora is in-scope binding. *Semantics and Pragmatics*, *1*(1), 1–42. DOI: https://doi.org/10.3765/sp.1.1

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. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Büring, D. (2004). Crossover situations. *Natural Language Semantics*, *12*(1), 23–62. DOI: https://doi.org/10.1023/B:NALS.0000011144.81075.a8

Carminati, M. N., Frazier, L., & Rayner, K. (2002). Bound variables and c-command. *Journal of Semantics*, *19*(1), 1–34. DOI: https://doi.org/10.1093/jos/19.1.1

Chow, W.-Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, *5*, 630. DOI: https://doi.org/10.3389/fpsyg.2014.00630

Cunnings, I., Patterson, C., & Felser, C. (2015). Structural constraints on pronoun binding and coreference: Evidence from eye movements during reading. *Frontiers in Psychology*, *6*. DOI: https://doi.org/10.3389/fpsyg.2015.00840

Elbourne, P. D. (2005). *Situations and individuals*. MIT Press.

Greene, S. B., Gerrig, R. J., Mckoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language, 33*(4), 511–526. DOI: https://doi.org/10.1006/jmla.1994.1024

Heim, I. (1990). E-type pronouns and donkey anaphora. *Linguistics and Philosophy*, *13*(2), 137–177. DOI: https://doi.org/10.1007/BF00630732

Hirschberg, J., & Ward, G. (1991). Accent and bound anaphora. *Cognitive Linguistics*, *2*(2), 101–122. DOI: https://doi.org/10.1515/cogl.1991.2.2.101

Kush, D. W. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing* [Doctoral dissertation, University of Maryland]. Digital Repository at the University of Maryland. https://drum.lib.umd.edu/handle/1903/14589

Kush, D., & Eik, R. (2019). Antecedent accessibility and exceptional covariation: Evidence from Norwegian donkey pronouns. *Glossa, 4*(1), 1–17. DOI: https://doi.org/10.5334/gjgl.930

Kush, D., Lidz, J., & Phillips, C. (2015). Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language, 82*, 18–40. DOI: https://doi.org/10.1016/j.jml.2015.02.003

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

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. DOI: https://doi.org/10.1016/j.tics.2006.08.007

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*(2), 111–123. DOI: https://doi.org/10.1023/A:1005184709695

Moulton, K. (2017). Retrieving antecedents in processing: Binder indices and φ-features. In W. G. Bennett, L. Hracs, & D. R. Storoshenko (Eds.), *Proceedings of the 35th West Coast Conference on Formal Linguistics* (pp. 30–40). Cascadilla Proceedings Project. http://www.lingref.com/cpp/wccfl/35/abstract3372.html

Moulton, K., & Han, C. (2018). C-command vs. scope: An experimental assessment of bound-variable pronouns. *Language, 94*(1), 191–219. DOI: https://doi.org/10.1353/lan.2018.0005

Poesio, M., & Zucchi, A. (1992). On telescoping. *Semantics and Linguistic Theory, 2*, 347–366. DOI: https://doi.org/10.3765/salt.v2i0.3034

Postal, P. (1966). On so-called "pronouns" in English. *Report on the Seventeenth Annual Round Table Meeting on Linguistics and Language Studies*, 177–206.

Reinhart, T. (1983). *Anaphora and semantic interpretation*. Routledge. DOI: https://doi.org/10.4324/9781315536965

Rothstein, S. (1995). Small clauses and copular constructions. In A. Cardinaletti, & M. T. Guasti (Eds.), *Small clauses* (Vol. 28, pp. 25–48). Brill. DOI: https://doi.org/10.1163/9780585492209_003

Safir, K. (2004). *The syntax of anaphora*. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195166132.001.0001

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