Journal of Memory and Language 82 (2015) 18–40

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Relation-sensitive retrieval: Evidence from bound variable pronouns

Dave Kush [a,*], Jeffrey Lidz [b], Colin Phillips [b]

[a] Haskins Laboratories, New Haven, CT 06511, USA
[b] Department of Linguistics, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 13 November 2013
revision received 16 February 2015

Keywords:
Pronouns
Binding
C-command
Memory retrieval
Sentence processing
Quantifiers

ABSTRACT

Formal grammatical theories make extensive use of syntactic relations (e.g. c-command, Reinhart, 1983) in the description of constraints on antecedent-anaphor dependencies. Recent research has motivated a model of processing that exploits a cue-based retrieval mechanism in content-addressable memory (e.g. Lewis, Vasishth, & Van Dyke, 2006) in which item-to-item syntactic relations such as c-command are difficult to use as retrieval cues. As such, the c-command constraints of formal grammars are predicted to be poorly implemented by the retrieval mechanism. We tested whether memory access mechanisms are able to exploit relational information by investigating the processing of *bound variable pronouns*, a form of anaphoric dependency that imposes a c-command restriction on antecedent-pronoun relations. A quantificational NP (QP, e.g., *no janitor*) must c-command a pronoun in order to bind it. We contrasted the retrieval of QPs with the retrieval of referential NPs (e.g. *the janitor*), which can co-refer with a pronoun in the absence of c-command. In three off-line judgment studies and two eye-tracking studies, we show that referential NPs are easily accessed as antecedents, irrespective of whether they c-command the pronoun, but that quantificational NPs are accessed as antecedents only when they c-command the pronoun. These results are unexpected under theories that hold that retrieval exclusively uses a limited set of content features as retrieval cues. Our results suggest either that memory access mechanisms can make use of relational information as a guide for retrieval, or that the set of features that is used to encode syntactic relations in memory must be enriched.

Introduction

Pronouns typically depend for their interpretation on antecedents in the previous linguistic and non-linguistic context. Accessing these antecedents in memory requires retrieval processes (Foraker & McElree, 2007; Gordon & Hendrick, 1998b; Sanford & Garrod, 2005). The relations between pronouns and their antecedents are also subject to numerous constraints, which have been extensively studied in linguistics and psycholinguistics. Pronoun resolution, therefore, provides a valuable test case for investigating the interplay of linguistic constraints and memory access mechanisms in language: by examining how constraints on pronoun antecedents guide antecedent retrieval processes, we can gain insight into how linguistic memory is encoded and navigated. In this study we focus on the resolution of so-called *bound variable pronouns*, because their standard linguistic analysis involves a

* Corresponding author at: Haskins Laboratories, 300 George St., Suite 900, New Haven, CT 06511, USA.
E-mail address: kush@haskins.yale.edu (D. Kush).

http://dx.doi.org/10.1016/j.jml.2015.02.003
0749-596X/

configurational constraint on antecedents that is not easily captured in otherwise well-motivated cue-based models of memory access.

Pronouns can have referential antecedents (1) or quantificational antecedents (2a,b).

(1) *The cyclist* was convinced that the spectators adored *him*.
(2a) *Every cyclist* thought that the spectators adored *him*.
(2b) *No cyclist* suspected that the spectators loathed *him*.

The choice of antecedent determines how the pronoun is interpreted. In (1) the pronoun *him* enters into a *co-reference* relation with its antecedent *the cyclist*; both expressions point to the same single individual in a discourse model (e.g., Büring, 2005). In (2a,b) the pronouns are said to be 'bound' by their respective antecedents *every cyclist* and *no cyclist*. These bound-variable pronouns do not refer to a single individual in the discourse model, but rather *co-vary in interpretation* with the quantified phrase (QP), which provides instructions on how to iterate through individuals in the discourse model.

Antecedent-pronoun relations are governed by two kinds of constraints: (i) morphological constraints, which enforce *feature-match* relations between the antecedent and the pronoun, and (ii) syntactic constraints, which determine an antecedent's eligibility based on its *relative structural position* to a pronoun. Morphological constraints apply to co-reference and binding relations alike; all pronouns must agree with their antecedents.

(3) {*The/No*} {*boy/*girl*} thought that the spectators adored *him*.

Certain syntactic constraints also apply uniformly to antecedent-pronoun relations, such as Principle B (Chomsky, 1981), which prohibits a pronoun from taking a prominent clause-mate antecedent.

(4) *{*The/No*} *boy* adored *him*.

Other syntactic constraints appear to target binding dependencies specifically. For example, the QP *no cyclist* cannot bind the pronoun *him* because it is embedded within a relative clause that does not contain the pronoun.

(5) The photographers [that **no cyclist* posed for] still had pictures of *him*.

Co-reference is not subject to the same restriction. A referential NP in the same position as the QP in (5) can readily serve as an antecedent for the pronoun.

(6) The photographers [that *the cyclist* posed for] still had pictures of *him*.

Thus, QP-pronoun binding relations are subject to a stringent positional constraint that does not influence NP-pronoun co-reference relations.

Many theorists have formalized this positional constraint in terms of c-command (Büring, 2005; Reinhart, 1983; among many others).[1] An item *X* c-commands another item, *Y*, if *Y* is contained within *X*'s sister in the syntactic tree (or is X's sister itself). For example, the quantificational phrases (QPs) in (2a,b) c-command the pronouns because they are contained within the verb phrase (VP) that is the QP's sister.

Relational constraints such as the c-command constraint on bound variable pronouns are particularly interesting for models of memory access in sentence processing because they pose a potential challenge for otherwise well-motivated models of retrieval. Popular cue-based models assume that retrieval makes use of intrinsic, item-specific features that are encoded during initial processing (e.g., Anderson, 1990; Gillund & Shiffrin, 1984; Lewis, Vasishth, & Van Dyke, 2006). These features can be drawn from an item's lexical entry (e.g., phrasal category, number, gender, and lexical semantics) or from its local syntactic context (e.g., grammatical role). It is straightforward to implement morphological feature-match constraints because a candidate antecedent's morphological features are item information (drawn directly from a noun's lexical entry). Accordingly, many studies report that a gender and/or number mismatch between an anaphor and a potential antecedent has immediate effects on early pronoun processing, consistent with the hypothesis that this information is used as a cue to guide retrieval (e.g., Badecker & Straub, 2002; Chow, Lewis, & Phillips, 2014; Ehrlich & Rayner, 1983; Garnham, Oakhill, Ehrlich, & Carreiras, 1995; Garrod & Terras, 2000; Gerrig, 1986; Nieuwland, 2014; Osterhout & Mobley, 1995).

C-command relations that encode the relative position of two distant items in a representation are difficult to encode as inherent features of individual items. First, a relation like *X c-commands Y* cannot be encoded through the use of a generalized ±c-command feature that marks the c-commander X as structurally prominent. Such an encoding scheme would fail to represent the crucial item-to-item configuration between X and Y. Encoding the relation on item X would therefore require a feature that made direct reference to Y (perhaps through use of a pointer as in [c-commands: Y]). Although features of this kind would be relatively easy to encode if X and Y were adjacent to one another, they present an encoding challenge as the distance between X and Y grows. When an incremental parser first encodes an item X, any subsequent item Y does not exist in the local syntactic context because Y has not yet been encountered. Encoding that X c-commands Y would require look-ahead, or prediction of Y. This might be possible in a narrow range of linguistic dependencies that are highly predictable, but that is less feasible in the case of pronouns, which are not, in general, predictable. Alternatively, encoding c-command relations would require that as each new item is introduced into the structure, all prior items that c-command that item are retroactively updated, which would impose a

[1] For purposes of the current article we adopt the standard view that the relational constraint on bound variable pronouns involves c-command. Some work has questioned whether c-command is the appropriate relational constraint (e.g., Barker, 2012), but there is little dispute over the notion that some kind of relational constraint is needed.

significant computational burden on the parser (see Kush, 2013). Encoding other item-to-item relations, such as serial order or precedence, faces a similar challenge. Moreover, there is evidence that serial order information is not used during direct-access retrieval (e.g., McElree, 2000; McElree, 2006; McElree, Foraker, & Dyer, 2003; cf. Öztekin & McElree, 2010). The implication is apparent: if information that encodes the relative position of two items cannot be used as a retrieval cue, c-command relations are also unlikely to be used as retrieval cues in accessing referential antecedents. In light of the centrality of c-command in the characterization of diverse linguistic constraints (e.g., Chomsky, 1981; Fiengo, 1977; Kayne, 1994), we therefore encounter a clear tension between well-motivated linguistic notions and well-motivated memory mechanisms.

Previous work has tested whether comprehenders pursue bound variable interpretations in real time sentence comprehension and has found that participants can rapidly compute a bound variable interpretation of a pronoun when that pronoun is c-commanded by a matching QP (e.g., Avrutin, 1994; Cunnings, Patterson, & Felser, 2014; Frazier & Clifton, 2000; Guo, Foley, Chien, Chiang, & Lust, 1996; Shapiro & Hestvik, 1995). This indicates that antecedent retrieval has access to QPs that are grammatically eligible to bind a pronoun, but it does not establish retrieval's sensitivity to relations such as c-command.

We aimed to determine whether initial antecedent retrieval also accesses non-c-commanding QPs whose structural relations to a pronoun make them ungrammatical as antecedents. We selected our test constructions with care to avoid certain construction types that could potentially obscure our ability to observe whether retrieval displays general sensitivity to syntactic relations.

First, there are some restricted syntactic environments where a true bound variable interpretation is acceptable in the absence of strict c-command (Barker, 2012; Büring, 2005). For example, QP possessors such as *each boy* in (7a) behave as if they have the c-command domain of the entire NP, i.e., *each boy's mother*, yielding acceptable bound variable interpretations. (See Gordon & Hendrick, 1998b for experimental validation of the acceptability of such constructions.)

(7a) Each boy's mother decided what he could do.
(7b) No boy's mother decided what he could do.

The acceptability of binding in (7a,b) suggests that the formulation of the constraint on bound-variable anaphora in terms of strict c-command may require some modification (Barker, 2012; Kush, 2013; May, 1977), but it does not undermine the inherently relational character of the constraint. Variable binding is still limited by the relative position of the QP to the pronoun, as evidenced by the fact that even a small change to the genitive relation in (7) yields a degraded bound variable interpretation.

(8) #The mother of no boy decided what he could do.

Because examples like (7) involve acceptable binding, they are not relevant to our primary question, which is whether initial retrieval respects the relational constraints that govern offline acceptability.

The second case that we avoided in our design involves situations where a pronoun is licensed by a QP, but does not yield a true bound variable interpretation. Such cases are well known in the semantics literature and are often referred to as cases of *Telescoping* (see, e.g., Anderssen, 2011; Poesio & Zucchi, 1992; Roberts, 1989). Failure to distinguish these constructions from cases of true variable binding could potentially lead to the mistaken conclusion that QP-pronoun binding is subject to less stringent relational constraints than it actually is. Fortunately, the constructions are generally restricted to certain types of quantifiers, and so they are easily avoided.

An instructive example of such a case can be found in an eye-tracking study by Carminati, Frazier, and Rayner (2002), who investigated how c-command affected the ease of processing a pronoun in the presence of a preceding noun phrase. The authors manipulated the referentiality of a potential antecedent noun phrase, e.g., *British soldier* in (9). The potential antecedent was either quantificational (9a,b) or referential (9c,d). They also manipulated whether the noun phrase c-commanded the pronoun, by placing the pronoun in either an embedded clause (c-command: 9a,c) or a conjoined clause (no c-command: 9b,d).

(9a) Every British soldier thought that *he* killed an enemy soldier.
(9b) Every British soldier aimed and then *he* killed an enemy soldier.
(9c) The British soldier thought that *he* killed an enemy soldier.
(9d) The British soldier aimed and then *he* killed an enemy soldier.

The authors found no evidence of increased difficulty in processing pronouns with non-c-commanding antecedents, even when the antecedent was quantificational, as in (9b). At first sight, these findings might be taken as evidence that non-c-commanding QPs can establish binding relations across conjuncts. However, (9b) is acceptable in offline judgments, and as Carminati and colleagues themselves noted, the QP-pronoun relation in sentences such as (9b) does not exhibit behavior characteristic of a true bound-variable dependency. The acceptability of true instances of variable binding does not depend on the choice of quantifier. Replacing *every* in (10a) with the negative quantifier *no* in (10b) yields an acceptable result. This is characteristic of true bound variable interpretations. In contrast, substituting a negative quantifier in the conjoined sentences in (10c,d) yields an unacceptable result.

(10a) *Every British soldier* thought that *he* killed an enemy.
(10b) *No British soldier* thought that *he* killed an enemy.
(10c) *Every British soldier* picked up his rifle and then *he* killed an enemy soldier.
(10d) **No British soldier* picked up his rifle and then *he* killed an enemy soldier.

Following Bosch (1983), Carminati and colleagues suggested that the pronoun in (9b) is not truly bound by the QP *Every British soldier*. Rather, it is interpreted as

co-referential with an *inferred antecedent* that serves as a stand-in for a representative participant in a generic (or stereotypical) statement. Through co-reference with this entity, the pronoun can be interpreted as co-variant with the *every* QP, thereby producing an interpretation very similar to one of true variable binding even though binding is not involved (see discussion of telescoping constructions in, e.g., Anderssen, 2011; Poesio & Zucchi, 1992). Negative quantifiers such as *no* do not license this inference, hence the unacceptability of (10d).

Under this interpretation, the comparison between (9a) and (9b) provides evidence that the retrieval of the inferred antecedent QP in (9b) imposes no greater processing demands than the retrieval of the true binder QP in (9a). Importantly, these findings do not address our main question: whether QPs that are judged to be unacceptable due to the c-command constraint are nevertheless accessed on-line as potential binders.

The discussion above underscores the importance of using test cases where a non-c-commanding QP is clearly unacceptable as a potential antecedent for a pronoun. In our studies we ensured that our test cases were suitable in two ways. First, we used offline judgments to establish that our test sentences do not permit bound variable interpretations. Second, we focused on the processing of negative quantifiers, such as *no*, to avoid the confound introduced by the alternative readings that are sometimes available to quantifiers like *every*.

Experiment 1: Distinguishing potential antecedents on the basis of c-command

Experiment 1 compared the effect of the position and the type of potential antecedents on the processing of pronouns. Specifically, we tested whether c-command imposes a constraint on the retrieval of a negative QP as an antecedent for a feature-matching pronoun. We compare the effect of c-command on the retrieval of QPs to its effect on the retrieval of referential NPs.

Materials

The experiment used a 2 × 2 factorial design, manipulating the factors Antecedent Type and Structure, as illustrated in Table 1. All test items consisted of a pair of clauses linked by either *but* or *when*. A singular pronoun appeared at the beginning of the second clause, and hence needed to find an antecedent in the first clause. The first conjunct began with a noun phrase (*Kathi*) that mismatched the gender of the pronoun and was followed by a second noun phrase (*janitor*) that matched the gender of the pronoun. This noun phrase was the *potential antecedent*. Potential antecedents were placed in an embedded clause to prevent any advantage in retrieval probability associated with first-mention, or with being the main subject of the sentence (Corbett & Chang, 1983; Gernsbacher & Hargreaves, 1988). A subsequent noun phrase in the first clause included a possessive pronoun (*his job*) that matched the gender of the antecedent. This possessive ensured that the antecedent was a highly accessible referent when readers reached the end of the first clause.

Table 1
Example items for Experiment 1. Slashes indicate regions of analysis used for Experiment 1c.

Condition	Example sentence
Quantificational-But	Kathi didn't think **any janitor** liked performing his/ custodial duties, /**but** he/ had to/ clean up messes/ left after prom anyway.
Quantificational-When	Kathi didn't think **any janitor** liked performing his/ custodial duties **w/hen** he/ had to/ clean up messes/ left after prom.
Referential-But	Kathi didn't think **the janitor** liked performing his/ custodial duties, /**but** he/ had to/ clean up messes/ left after prom anyway.
Referential-When	Kathi didn't think **the janitor** liked performing his/ custodial duties **w/hen** he/ had to/ clean up messes/ left after prom.

The factor Structure manipulated the attachment height of the clause containing the critical pronoun. The pronoun was either contained in a temporal adverbial clause introduced by *when*, which attaches to the embedded VP of the first conjunct, or inside a clause introduced by *but*, which attaches to the root of the sentence. The coordinator *but* was chosen over *and*, because *but* is less ambiguous in its attachment options. By manipulating the attachment height of the clause containing the pronoun, we manipulated whether a c-command relation held between the pronoun and the feature-matching NP, without changing the position of the feature-matching NP. In the *When* conditions the pronoun was c-commanded by the antecedent (see Fig. 1). In the *But* conditions the antecedent did not c-command the pronoun (Fig. 2).

The conditions compared the effect of c-command as manipulated by Structure on the accessibility of feature-matching *Referential* and *Quantificational* noun phrases in matched structural positions. This structural manipulation affects the ability of the quantificational noun phrases to bind the pronoun, but it does not affect the referential phrase's ability to serve as an antecedent because co-reference is not predicated on c-command. The factor Antecedent Type manipulated the determiner of the feature-matching potential antecedent. In the *Referential* conditions the determiner was definite, and in the *Quantificational* conditions the determiner was the quantifier *any*, the negative polarity item (NPI) counterpart of the quantifier *no* (Ladusaw, 1980). As stated above, we chose to use a negative quantifier like *any* because negative quantifiers such as *any* or *no* exhibit greater restrictions in their scope-taking abilities (see e.g., Beghelli & Stowell, 1997) and they do not license the referential interpretation that confounded the results of Carminati et al. (2002). The quantifier *any*, with a preceding negation to license it, was chosen over the quantifier *no* in order to maximize the naturalness of the example sentences.

We predicted that the feature-matching referential NP should always be accessible to retrieval, irrespective of

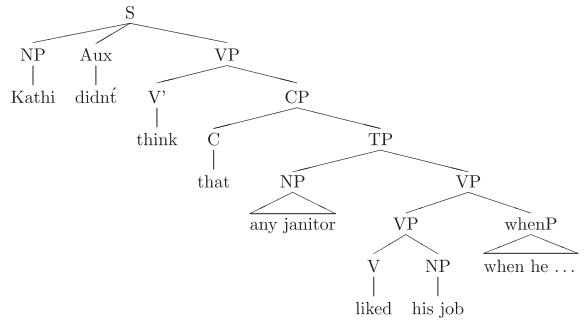

Fig. 1. Schematic of the syntactic structure of *when*-conditions in Experiment 1.

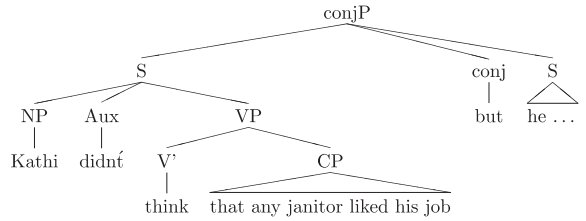

Fig. 2. Schematic of the syntactic structure of *but*-conditions in Experiment 1.

the connective, because c-command is not a pre-condition on coreference between a referential NP and a pronoun. Our experiment tested retrieval's sensitivity to relations by measuring the impact that the choice of connective exerts on pronominal processing when the matching NP is quantificational. In the *Quantificational-When* condition, the feature-matching QP c-commands, and can therefore bind, the pronoun. In contrast, the QP fails to c-command the pronoun in the *Quantificational-But* condition, so it cannot act as a binder according to the constraint.

If antecedent retrieval is not sensitive to relational information, we expect the *but/when* manipulation to have no effect on early processing of the pronoun. However, if antecedent retrieval is sensitive to relational information, then we predict an Antecedent Type × Structure interaction. Antecedent retrieval should be able to readily access the matching, c-commanding QP in the *Quantificational-When* condition, and difficulty is expected in the *Quantificational-But* condition.

Experiment 1a: Acceptability rating task (offline)

As a first step in testing the effect of structure on resolving bound variable pronouns we sought to verify that the non-c-commanding QPs are not acceptable antecedents for the critical pronoun in our test sentences. We used two offline measures: simple acceptability ratings and a paraphrase selection task as a measure of interpretation. Because we were specifically interested in assessing whether antecedent retrieval respects distinctions made by the grammar, it is important to establish that our test sentences reflect cases where a grammatical c-command constraint is in effect.

Participants

16 participants (mean age = 34.9, 7 male) were recruited through the Amazon Mechanical Turk (AMT) marketplace and paid $4.00 for their participation (see Sprouse, 2011 for discussion of the reliability of AMT-collected judgment data). Participants were qualified to participate based on US-internal IP addresses and passing a native-speaker screening test, consisting of a short series of acceptability judgments (unrelated to the current study) that prior testing had shown to effectively distinguish native and non-native speakers of English.

Procedure

Presentation used the IBEX Farm internet-based, experimental presentation platform (Drummond, 2011). Sentences were presented one at a time centered on the screen. A 7-point acceptability scale was presented as an array of numbered boxes below each sentence, with end-points marked 'bad' (1), and 'good' (7). Participants were instructed to rate sentences on the scale, with a rating of '1' corresponding to 'bad', 'unacceptable' or 'doesn't make sense', and a rating of '7' corresponding to 'good', 'totally acceptable', or 'easy to understand'. Participants were not explicitly instructed to attend to pronouns, but practice items illustrated cases where intra-sentential coreference and binding were unavailable.

Analysis

Acceptability ratings from each participant were *z*-scored using both test and filler items prior to analysis (Schütze & Sprouse, 2014). Statistical analysis used linear mixed-effect models (LMEMs) with maximal random-effects structures. Each model included simple difference sum-coded fixed effects of Antecedent Type (whether the feature-matching NP was *Referential*, or *Quantificational*), Structure (whether the critical pronoun was preceded by *but*, or *when*), and their interaction and random intercepts for participants and items (Baayen, Davidson, & Bates, 2008). Random slopes were also included for all fixed effects and their interaction when models converged (Barr, Levy, Scheepers, & Tily, 2013). If the model failed to converge with a maximal random-effects structure, random slopes for items were removed. A fixed effect was considered significant if its absolute *t*-value was greater than 2, which indicates that its 95% confidence interval did not include 0 (Gelman & Hill, 2007). Reported coefficients whose absolute *t*-value was greater than 2 are significant at the $p < .05$ level. We adopt the assumption that a reported coefficient is marginally significant ($p < .10$) if the absolute value of its *t*-value is greater than 1.65 (based on the 90% confidence interval).

Materials

24 sets of four experimental items were distributed across four lists according to a Latin Square, and were interspersed in a pseudo-randomized order among 64 fillers, resulting in lists of 88 items per participant, composed

of roughly 60% acceptable and 40% unacceptable sentences. Eighteen of the unacceptable fillers contained errors such as unlicensed negative polarity items, subject-verb agreement errors, or were missing function words such as *the*. Eight of the unacceptable fillers contained QPs and pronouns that could not be bound by them (due either to number mismatch or to structural constraints). These 8 filler sentences, referred to below as *Bad Binding* sentences, were included to provide a baseline against which to compare the acceptability of the *Quantificational-But* conditions. Six of the fillers contained acceptable examples of structurally licensed variable binding, and served as an independent assessment of participants' willingness to accept a bound variable interpretation of the pronoun when it was the only viable option. These are referred to below as *Good Binding* sentences. Four additional unacceptable sentences featured an infelicitous pronoun inside an adverbial clause introduced by *when*, to prevent participants from developing superficial tendencies to rate all sentences containing *when* as acceptable due to its frequency in the test-items. Similarly, four sentences featuring an infelicitous pronoun and the connectives *but* or *and* were included for the same reason. 24 additional acceptable fillers were also included.

Results

Mean acceptability judgments, both raw and *z*-scored, are provided in Table 2. All analyses reported here were conducted on *z*-scored ratings. A reliable effect of Antecedent Type was observed ($\beta = -0.60$, s.e. = .08, $t = -7.89$), driven by higher scores in the *Referential* conditions than in the *Quantificational* conditions. A marginally significant effect of Structure was observed ($\beta = -0.14$, s.e. = 0.07, $t = -1.88$), due to higher average acceptability scores in the *When* conditions. The Antecedent Type × Structure interaction was significant ($\beta = -0.37$, s.e. = .15, $t = -2.47$), with Quantificational-But conditions receiving the lowest acceptability scores overall. *Quantificational-When* sentences were rated reliably more acceptable than *Quantificational-But* sentences ($t = -2.86$). No comparable pairwise difference was observed between the Referential conditions ($t < 1$).

Because control sentences were not factorially manipulated we did not perform statistical comparisons of the ratings. *Good-Binding* control sentences received numerically higher acceptability ratings than did *Bad-Binding* sentences and *Bad-Filler* sentences. *Bad-Binding* sentences were also rated as more acceptable on average than *Bad-Filler* sentences.

Table 2
Average raw and *z*-scored acceptability ratings for items in Experiment 1a. Standard errors in parentheses.

Condition	Average raw rating	Average *z*-score
Referential-But	5.77 (.16)	0.69 (.07)
Referential-When	5.67 (.14)	0.64 (.07)
Quantificational-But	4.18 (.20)	−0.11 (.09)
Quantificational-When	4.96 (.18)	0.29 (.08)
Filler: Good-Binding	5.82 (.16)	0.72 (.08)
Filler: Bad-Binding	4.28 (.17)	−0.03 (.08)
Filler: Bad-Filler	3.05 (.11)	−0.62 (.05)

Discussion

The acceptability judgment experiment sought to experimentally test the generalization that c-command affects the acceptability of a QP, but not a referential NP, as an antecedent for a pronoun. The qualitative pattern of acceptability ratings is consistent with the generalization from the linguistics literature. C-command had a clear effect on the accessibility of QP antecedents, but not on participants' ability to establish a co-reference relation between a referential antecedent and a critical pronoun.

The results are sufficient to show that there is an isolable effect of c-command that can be investigated in online measures. The results are more equivocal on whether c-command categorically blocks the non-c-commanding QP and the pronoun from entering into a binding relation. The relatively high acceptability scores in the *Quantificational-But* condition are unexpected if a pronoun that lacks an overt antecedent should have induced unacceptability (e.g., Gordon & Hendrick, 1998b). The acceptability scores are consistent with two interpretations. First, they might suggest that c-command influences the acceptability of a QP-pronoun binding dependency, but that it is not the sole determinant of establishing such a relation. On this interpretation, the QP and the pronoun would be linked in the absence of c-command, but at a cost. The second interpretation is that the pronoun was interpreted as disjoint from the QP in accordance with the grammatical generalization, but that the accommodation of an antecedent-less pronoun did not result in as dramatic a reduction in acceptability as was initially predicted. This second interpretation is consistent with findings from the similar acceptability ratings between the *Quantificational-But* and the *Bad-Binding* conditions. *Bad-Binding* sentences contained no feature-matching antecedent for the pronoun, so acceptability ratings in this condition may serve as an indicator of the cost of accommodating a pronoun without an antecedent. In order to distinguish between these possibilities we conducted a test of participants' interpretations of these sentences.

Experiment 1b: Sentence judgment study (Offline)

In order to test whether participants entertain a binding dependency between the non-c-commanding QP and the pronoun in Experiment 1 we conducted a forced-choice paraphrase task.

Participants

22 participants were recruited through Amazon Mechanical Turk and paid $3.50 for their participation.

Materials and procedure

The paraphrase task used the same target sentences as Experiment 1a. A target sentence was drawn from the

items from the 2 × 2 design above and paired with two corresponding response sentences that represented possible paraphrases of the test sentence. These response sentences were designed to be either consistent with a *single-individual* interpretation of the pronoun in the test sentence (response sentences marked *SI*), or a *quantificational* interpretation of the pronoun (marked *Q*). Sentences consistent with a single-individual interpretation used an existential construction to assert the existence of an individual who performed the action described by the post-pronominal VP. Sentences consistent with a quantificational interpretation used a paraphrase of the relevant portion of the test sentence using the quantifier *every*, instead of *any*. Predicates in the response sentences differed minimally based on the semantics of the test sentence.

(11) Example But Sentence
Kathi didn't think any janitor enjoyed performing his custodial duties, but he had to clean up the messes left after prom anyway.
SI. There was someone who had to clean up after prom.
Q. Every janitor had to clean up after prom.

(12) Example When Sentence
Kathi didn't think any janitor enjoyed his custodial duties when he had to clean up the messes left after prom.
SI. There was someone who disliked having to clean up after prom.
Q. Every janitor disliked having to clean up after prom.

Items were distributed in a Latin Square design across four lists, and presented in a pseudo-randomized order. On each trial participants read the sentence triplet and chose the paraphrase that best matched their interpretation of the test sentence. Response sentences were presented as a numbered list in randomized order. Test sentences were interspersed randomly among 28 filler triplets, which contained a pronoun consistent with a single-individual (*Filler-Single*) or quantificational (*Filler-Quantificational*) interpretation. *Filler-Single* sentences often contained QPs that could not bind the pronoun due to feature-mismatch (e.g., Test Sentence: *The programmer tried to speak with no girls, so* ***she*** *stopped trying to talk with him.* SI Paraphrase: *There was someone that stopped trying to talk to the programmer.* Q Paraphrase: *Every girl stopped trying to talk to the programmer.*). *Filler-Quantificational* sentences contained a matching QP that c-commanded the pronoun (e.g., *Every seamstress that the tailor made dresses with said she could have designed something better.* SI Paraphrase: *There was someone that thought her skills were superior to the tailor's.* Q Paraphrase: *Every seamstress thought that her skills were superior to the tailor's.*).

Participants were encouraged to answer accurately and received feedback on their accuracy on both fillers and *Referential* test sentences. This step was taken to ensure that participants were attending to the task and interpreting the sentences. Triplets containing *Quantificational* test sentences were not coded as having correct responses, so participants received no error message on these trials regardless of their input. This was done to avoid influencing participants' response preferences on *Quantificational* items.

Analysis

Data from five participants who scored lower than 70% accuracy on either the filler questions or *the Filler-Single* test sentences were excluded from analysis. Statistical analysis used logistic-mixed effect models with the structure described in Experiment 1a.

Results

The proportion of trials, by condition, on which participants chose the sentence corresponding to the quantificational interpretation of the continuation (e.g. *every janitor*) is shown in Table 3.

Statistical analysis revealed a significant main effect of Structure ($\beta = -2.24$, s.e. = 0.45, $z = -5.000$). Participants chose Quantificational paraphrases more often in *When* conditions than in *But* conditions. A main effect of Antecedent Type was also significant ($\beta = -4.25$, s.e. = 0.49, $z = -8.626$), driven by an increased proportion of quantificational responses in *Quantificational* conditions relative to *Referential* conditions. The crucial Structure × Antecedent Type interaction was also observed ($\beta = -2.93$, s.e. = 0.88, $z = -3.307$). This reflected a higher proportion of Quantificational responses in the *Quantificational-When* condition. The Quantificational response was chosen more often in the *Quantificational-When* condition than in the *Quantificational-But* condition ($\beta = 4.31$, s.e. = 0.66, $z = 6.495$), the *Referential-But* condition ($\beta = 8.41$, s.e. = 1.87, $z = 4.486$), and the *Referential-When* condition ($\beta = 5.69$, s.e. = 0.92, $z = 6.212$). Pairwise comparisons also revealed that Quantificational responses were chosen more often in the *Quantificational-But* condition than in the *Referential-When* ($\beta = 2.06$, s.e. = 0.52, $z = 3.936$) and the *Referential-But* conditions ($\beta = 3.23$, s.e. = 0.76, $z = 4.241$).

Quantificational paraphrases were chosen more often in the *Filler-Quantificational* condition than in the *Filler-Single* conditions ($\beta = -3.40$, s.e. = 0.26, $z = -12.960$).

Discussion

Experiment 1b probed participants' preferred interpretations for pronouns in test sentences by having them select paraphrases that were consistent with either bound or referential readings of the pronouns. The c-command relation between a matching potential antecedent and the pronoun was manipulated, as was the

Table 3
Proportion of quantificational paraphrases chosen by participants in Experiment 1b.

	When	But
Quantificational	0.91	0.28
Referential	0.03	0.06
	Quantificational	Single
Filler	0.91	0.24

quantificational status of the potential antecedent. When the potential antecedent was referential, the pronoun could be coreferential with the potential antecedent. When the potential antecedent was a QP, an acceptable binding relationship was only possible if the QP c-commanded the pronoun. We measured participants' consideration of the bound interpretation in both conditions as the proportion of trials where they chose the quantificational paraphrase.

In the filler conditions, participants reliably picked the correct interpretation. In the *Referential* conditions, participants consistently chose the single-individual reading. This suggests that participants understood the task, and did not have difficulty with the existential construction used in single-individual responses.

Participants preferred to resolve the pronoun in *Quantificational-But* sentences as referring to an extra-sentential individual and not as a bound pronoun. On around 73% of trials participants favored coercing an unknown referent for resolution of the pronoun, rather than considering a grammatically illicit bound reading with the intra-sentential QP. Individual participant analyses show that most participants chose the quantificational interpretation for *Quantificational-But* trials at very low rates. Twelve of the seventeen participants chose the quantificational interpretation on one third or fewer trials. Five participants never chose the quantificational response on *Quantificational-But* trials, four chose the quantificational response in one out of six trials, and three chose the response on two out of six trials. The remaining five participants chose the quantificational paraphrase on greater than or equal to half of all *Quantificational-But* trials (range 0.50–1.00).

It is unclear what factors influenced participants to choose the ungrammatical quantificational paraphrase in the *Quantificational-But* condition more often than in either of the *Referential* conditions. The data are consistent with a number of possible interpretations that we cannot distinguish between. First, the data might indicate that participants considered bound readings in violation of the relational constraint to varying degrees. Alternatively, it is possible that participants did not consider the illicit bound reading and that the choice of the quantificational paraphrase reflects response pressures inherent in the forced-choice task. Extra-grammatical factors may have interacted in the task to render quantificational paraphrases more tempting in the *Quantificational-But* condition after participants encountered difficulty with interpreting the pronoun. For example, participants may have been more likely to choose a quantificational paraphrase when a QP was present, regardless of whether the QP could bind the pronoun. It may be that participants were more likely to judge the existential SI paraphrase as infelicitous when the pronoun lacked a clear antecedent. We also observe that some of the quantificational responses could be judged true based on world knowledge without reference to the test sentence or could be read as generic statements (e.g., Every fan is upset by watching better teams beat his team). These items may have acted as lures to participants who were uncertain how to interpret the antecedent-less pronoun. These extra-grammatical factors are less likely to have had an effect on *Referential* conditions, where participants did not face difficulty interpreting the pronoun.

Although the results are somewhat equivocal on the strength of the c-command effect, the forced-choice interpretation study nevertheless demonstrates that the c-command manipulation reliably influences participants' willingness to consider bound variable readings of the pronouns in our test sentences.

Experiment 1c. Eye-tracking while reading

Having established participants' offline sensitivity to the grammatical generalization that a QP cannot antecede a pronoun it does not c-command, we tested whether this offline sensitivity maps transparently onto real-time behavior. In an eye-tracking study, we investigated whether a feature-matching QP that does not c-command the pronoun is initially retrieved as a potential antecedent for a pronoun.

Participants

48 participants were recruited from the University of Maryland community (28 females, mean age 21.2). Participants received course credit or $10 for an hour of their time. All participants had normal, or corrected-to-normal vision, and were self-reported native speakers of English.

Materials and procedure

The 24 experimental item sets exemplified in Table 1 were distributed into 4 lists in a Latin Square design. Each list also contained 40 sentences from another experiment on pronoun resolution and 40 filler sentences for a total of 104 sentences per participant. The order of each list was pseudo-randomized such that no two experimental sentences were presented in succession. Sentences were presented in 12-point Courier font. The maximum number of characters allowed on a single line on the visual display was 142 characters, and all sentences in the experiment fit on one line. Eye movements were recorded using an Eyelink 1000 tower-mount eye-tracker, which sampled eye-movements at 1000 Hz. Participants had binocular vision while movements were measured, but only the right eye was tracked. The tower was 32 inches from the visual display, giving participants approximately 5 characters per degree of visual angle.

Before beginning the experiment, participants were familiarized with the apparatus and given four practice trials. While seated, participants' heads were immobilized using a chin rest and forehead restraint that was adjusted for comfort. Before the experiment, and whenever necessary throughout the experiment, the experimenter calibrated the eye-tracker with a 9-point display to ensure an accurate record of eye-movements across the screen. Participants began each experimental trial by fixating on a marker at the beginning of the sentence, triggering display of the entire test sentence. Participants terminated

the presentation of sentences via button-press on a response pad, which triggered presentation of a yes/no comprehension question. Participants were allowed to take breaks at their discretion throughout the experiment. Following each break, participants were recalibrated to ensure accurate measurement.

Data analysis

Test sentences were divided into 5 regions of interest, as indicated in Table 1. The sentence-initial region comprised all words from the beginning of the sentence to two words before the manipulated connective. The pre-critical region comprised the two pre-connective words. The critical pronoun region contained the last three letters of the connective and the critical pronoun. Previous studies (e.g., van Gompel & Majid, 2004) have defined a critical pronoun region in this manner to increase the size of the region in order to avoid excessive skipping (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994), as well as to account for potential parafoveal processing of the pronoun. The post-pronoun region contained the first two words of the verb phrase that followed the pronoun. The pronoun + 2 region contained the next two or three words in the verb phrase. We present data from four regions: pre-critical, critical pronoun, post-pronoun and pronoun + 2 regions. Average skipping rates across conditions in these regions were 8%, 6%, 9%, and 7% with a range of 2–14%.

For each region of interest we report four measures: first-pass, right-bound, second-pass, and total reading times. First pass reading time is calculated by summing all fixations in a region of interest after participants first enter the region until the first saccade out of that region (either to the right or the left). Right-bound reading time is the sum of all fixations in a region beginning when the region is first entered from the left to when it is first exited to the right. Right-bound times includes fixations that occur if a participant makes a leftward regression, and then re-enters the region. First-pass times therefore form a subset of right-bound times. Second-pass times are calculated by summing all fixations that occur in a region after a participant has first exited the region to the right. For second-pass time measures, trials on which a region is not re-fixated contribute a value of 0 ms to the cell mean. Total times sum over all fixations in a particular region of interest, including first pass reading time and any time spent rereading the region. Fixations under 80 ms were discarded (Rayner & Pollatsek, 1989). Extreme outliers, which were rejected by visual inspection, comprised less than 1% of the data across all measures and regions. Statistical analysis was carried out for each measure and region using LMEMs with the structure described in Experiment 1a.

Results

Table 4 provides mean raw reading times by measure and region. Table 5 provides a summary of statistical effects. Fig. 3 plots three interaction effects indicative of c-command sensitivity, two of which occur in the post-pronoun region and the other of which occurs in the pre-pronoun region.

Pre-critical region

There were no significant effects of Structure, Antecedent Type, or their interaction in either first-pass or right-bound reading times. In total times the *When* conditions were read more slowly on average than the *But* conditions ($t = -2.103$). In second-pass and total times mean reading times in the *Quantificational-When* condition were numerically higher than in any other condition. This Antecedent Type × Structure interaction was marginally significant in second-pass times ($t = -1.774$), but achieved significance in total times ($t = -2.045$). Mean total times in the *Quantificational-When* condition were significantly shorter than in the *Quantificational-But* condition ($\beta = -117.24$, s.e = 41.78, $t = -2.806$). Total reading times did not differ between the two *Referential* conditions ($t < 1$). Post-hoc analyses revealed that this interaction was largely driven by regressions from the post-pronoun region or later.

Table 4
Mean raw reading times by measure and region for Experiment 1c. Standard errors in parentheses.

		Pre-pronoun	Pronoun	Post-pronoun	Pronoun + 2
First-pass	*Quant-But*	430 (15)	364 (12)	345 (14)	405 (12)
	Quant-When	455 (15)	398 (13)	306 (9)	404 (14)
	Refer-But	455 (16)	377 (13)	309 (11)	389 (14)
	Refer-When	447 (15)	412 (13)	329 (10)	383 (13)
Right-bound	*Quant-But*	569 (19)	405 (16)	382 (16)	434 (13)
	Quant-When	570 (17)	445 (15)	334 (10)	449 (15)
	Refer-But	587 (19)	410 (17)	359 (14)	432 (15)
	Refer-When	545 (17)	449 (15)	364 (11)	424 (14)
Second-pass	*Quant-But*	280 (23)	227 (17)	165 (15)	194 (18)
	Quant-When	366 (32)	284 (25)	192 (17)	258 (22)
	Refer-But	291 (28)	239 (20)	186 (17)	207 (18)
	Refer-When	285 (23)	246 (22)	162 (15)	229 (19)
Total time	*Quant-But*	720 (27)	576 (22)	576 (20)	588 (21)
	Quant-When	837 (36)	673 (26)	673 (21)	659 (26)
	Refer-But	736 (27)	603 (24)	603 (22)	600 (23)
	Refer-When	752 (38)	652 (24)	652 (19)	612 (23)

Table 5
Summary of results of mixed effects models by region and measure for reading times in Experiment 1c. Results correspond to model estimates of each fixed effect's coefficient. Random intercepts were included for subjects and items, as were by-subject and by-item random slopes for all fixed effects and their interactions. Significant coefficients ($|t| > 2$) are in bold.

		Pre-pronoun		Pronoun		Post-pronoun		Pronoun + 2	
		Estimate (s.e.)	*t*-value	Estimate (s.e.)	*t*-value	Estimate (s.e.)	*t*-value	Estimate (s.e.)	*t*-value
First-pass	Intercept	**566.90 (31.2)**	**18.179**	**384.00 (18.5)**	**20.780**	**316.39 (19.2)**	**16.514**	**389.37 (22.2)**	**17.527**
	Structure	18.91 (19.1)	0.993	**−34.97 (13.9)**	**−2.515**	7.05 (13.8)	0.513	6.89 (13.7)	0.504
	Antecedent	3.40 (16.6)	0.205	−11.85 (11.4)	−1.043	5.31 (12.8)	0.416	18.75 (14.1)	1.330
	Struct × Anteced.	−45.19 (31.9)	−1.415	3.64 (23.1)	0.158	**50.87 (21.4)**	**2.372**	−2.05 (24.2)	−0.085
Right-bound	Intercept	**566.90 (31.2)**	**18.179**	**422.58 (23.1)**	**18.303**	**352.38 (22.5)**	**15.639**	**428.74 (25.1)**	**17.087**
	Structure	18.91 (19.1)	0.993	**−42.13 (15.6)**	**−2.700**	16.53 (16.3)	1.011	3.50 (13.2)	0.265
	Antecedent	3.40 (16.6)	0.205	−0.40 (14.4)	−0.028	−3.20 (12.7)	−0.251	10.85 (14.5)	0.746
	Struct × Anteced.	−45.19 (32.0)	−1.415	2.46 (28.1)	0.088	**49.87 (22.6)**	**2.207**	−26.31 (28.7)	−0.918
Second-pass	Intercept	**305.59 (37.5)**	**8.143**	**248.80 (26.7)**	**9.303**	**176.02 (20.0)**	**8.784**	**221.77 (26.3)**	**8.436**
	Structure	−39.92 (30.0)	−1.331	−31.84 (34.6)	−0.921	−0.95 (18.4)	−0.052	−42.58 (29.5)	−1.443
	Antecedent	34.14 (26.7)	1.278	12.42 (21.0)	0.590	3.85 (20.2)	0.191	7.12 (18.8)	0.380
	Struct × Anteced.	−92.06 (51.9)	−1.774	−48.73 (47.7)	−1.021	−52.27 (35.9)	−1.457	−40.78 (38.5)	−1.060
Total time	Intercept	**758.58 (49.1)**	**15.461**	**622.70 (34.7)**	**17.957**	**489.19 (30.9)**	**15.851**	**491.96 (32.0)**	**15.384**
	Structure	**−65.53 (31.2)**	**−2.103**	**−71.16 (33.7)**	**−2.109**	15.07 (24.4)	0.618	15.64 (25.0)	0.624
	Antecedent	32.20 (27.5)	1.171	−0.28 (21.5)	−0.013	4.02 (18.9)	0.213	3.48 (18.3)	0.190
	Struct × Anteced.	**−105.42 (51.6)**	**−2.045**	−50.39 (50.9)	−0.991	12.70 (41.3)	0.307	1.88 (42.3)	0.044

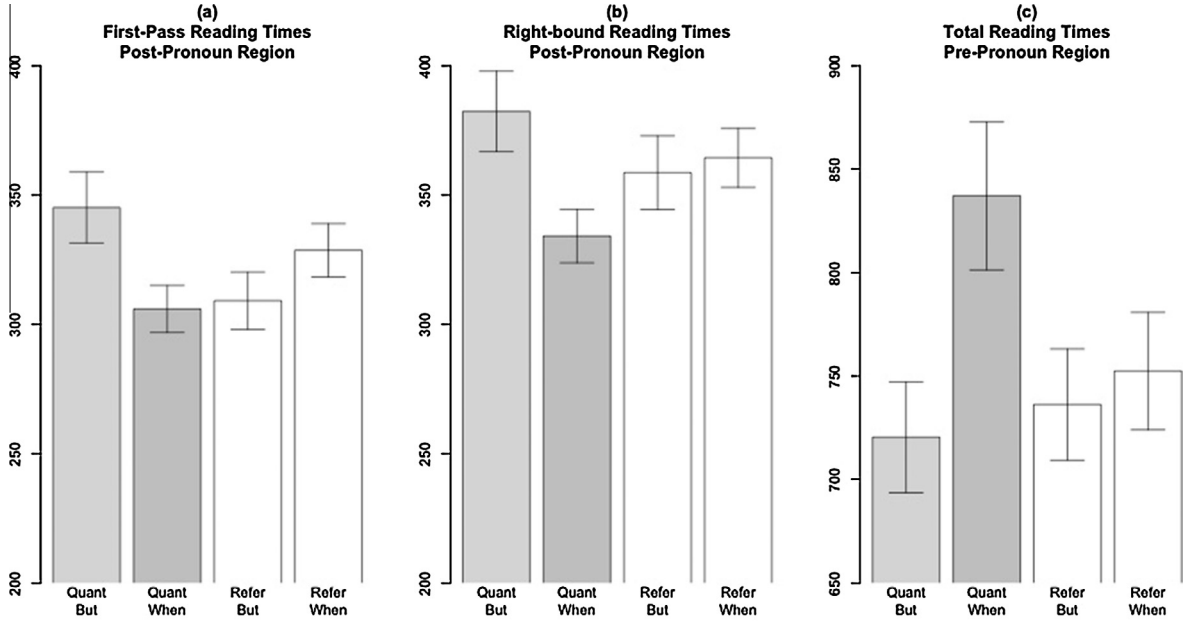

Fig. 3. (a) First-pass, (b) right-bound reading times for the post-pronoun region in Experiment 1c, and (c) total reading times in the pre-pronoun region. Error bars indicate one standard error of the mean.

Critical pronoun region

At the critical pronoun the *When* conditions were read more slowly than the *But* conditions in first-pass ($t = -2.312$), right-bound ($t = -2.492$), and total reading times ($t = -2.226$). No other effects were reliable in this region.

Post-pronoun region

There were no main effects in the region immediately following the pronoun, but reading times were characterized by an Antecedent Type × Structure interaction. This interaction was significant in first-pass ($t = 2.314$). First-pass reading times were longer on average in the *Quantificational-But* condition than in the *Quantificational-When* condition, but this difference was only marginally significant ($\beta = 31.14$, s.e. = 17.19, $t = 1.811$). The corresponding pairwise comparison between *Referential* conditions was not significant ($|t| < 1$). The same pattern manifested as a significant Antecedent Type × Structure interaction in right-bound times ($t = 2.207$). Right-bound reading times in the *Quantificational-But* condition were significantly longer than in the *Quantificational-When*

condition ($\beta = 41.18$, s.e. = 19.90, $t = 2.069$), but no differences were observed between the *Referential* conditions ($|t| < 1$).

Pronoun + 2 region

In the penultimate region there were no significant main effects or interactions in any measure.

Discussion

The study tested whether antecedent retrieval can distinguish between grammatically appropriate and grammatically inappropriate NPs when the c-command relation between the NP and the pronoun is the primary determinant of antecedent eligibility. The study compared the accessibility of a QP that c-commanded a pronoun (and was therefore a grammatical antecedent for that pronoun) with that of a QP that did not c-command the pronoun (and hence was not a grammatical antecedent). The study varied the structural relation between the antecedent and the pronoun while holding linear distance constant by manipulating a connective that controlled the attachment height of the clause containing the pronoun. Unlike quantificational binding, coreference between a referential NP and a pronoun is not subject to a c-command constraint. Conditions containing referential NPs in place of QPs provided a comparison point.

Offline measures supported the grammatical generalization that c-command between a QP and a pronoun is a primary factor in determining the acceptability of bound variable interpretations. Participants rated sentences in which a pronoun was c-commanded by a feature-matching QP as more acceptable than sentences in which no c-command relation obtained. This preference for a c-commanding antecedent was not observed in minimally different sentences in which the feature-matching phrase was a referential NP. A second study showed that on most trials participants chose to interpret a pronoun as bound by a QP only when the QP c-commanded the pronoun. On a small number of trials participants selected a paraphrase that reflected a bound variable interpretation that violated the c-command constraint. It is unclear whether these responses reflected genuine acceptance of variable binding, or whether they were a consequence of the forced-choice paraphrase task, where participants may have found neither alternative fully acceptable.

The first effects observed in the eye tracking record occurred immediately at the pronoun region. First-pass and right-bound reading times were shorter on average in *But* conditions than in *When* conditions. We attribute this reading time difference to the lexical difference between the conditions, since the pronoun region contained the last three characters of the manipulated connective, and to the fact that the lexical difference entailed an attachment difference.

The pattern of results at the post-pronoun region supports the contention that the c-command constraint on bound variable pronouns influences antecedent retrieval. The c-command manipulation had no reliable effect on the Referential conditions, in either first-pass or right-bound reading times. C-command had a clear effect on the Quantificational conditions in the same region. The *Quantificational-When* condition was read more quickly than its *Quantificational-But* counterpart suggesting that comprehenders had no difficulty accessing a c-commanding quantificational antecedent, but it did encounter difficulty processing the pronoun when the feature-matching QP did not c-command it.

The processing difficulty in the *Quantificational-But* condition is clear evidence that c-command impacts retrieval's ability to access a matching QP. What it does not determine, however, is the ultimate source of the difficulty. There are a few possible options. The difficulty in the *Quantificational-But* condition could reflect outright retrieval failure – because there are no matching NPs in the sentence, retrieval would not return any potential antecedents. On this account, the difficulty indexes the effort required to coerce a referent for the unheralded pronoun (e.g., Gerrig, 1986; Greene, Gerrig, McKoon, & Ratcliff, 1994). Another option is that retrieval erroneously returns *Kathi* as the potential antecedent for the pronoun instead of the non-c-commanding QP, which results in difficulty when the gender mismatch is detected. This possibility is also consistent with the interpretation that retrieval is sensitive to the c-command relation between the QP and the pronoun because the explanation requires that retrieval favors *Kathi* over the matching QP in the *Quantificational-But* condition alone. The third possible interpretation is that the longer reading times index partial-match retrieval interference instead of failure to retrieve the non-c-commanding QP altogether. The inappropriate QP might be (partially) activated via feature-match with the gender features of the critical pronoun, but it would not achieve the same level of activation as a QP that matched gender features and c-commanded the pronoun. This decreased activation would result in slower overall retrieval in comparison to retrieval of the matching, c-commanding QP on the assumption that retrieval latencies are inversely proportional to degree of feature-match (see, e.g. Anderson, 1990; Lewis & Vasishth, 2005). Once again, this option still requires retrieval to make a distinction between matching QPs on the basis of c-command.

Before concluding discussion of Experiment 1c, we mention one additional effect that provides evidence that may indirectly provide evidence that c-command influences processing. In the pre-pronoun region second-pass and total reading times were longer in the *Quantificational-When* condition, i.e. the only condition in which variable binding is grammatically possible, than in the remaining three conditions. Although we refrain from extensive speculation on the origins of this difference, we propose that it may indicate that the interpretation that comprehenders settle on in the *Quantificational-When* condition is qualitatively different than the interpretations pursued in the other conditions.

The results of Experiment 1 indicate that c-command has an immediate effect on the retrieval of QPs for variable binding. Processing of a pronoun is more effortful when the only feature-matching phrase in the discourse is a non-c-commanding QP than when a feature-matching QP c-commands the pronoun. Although the study shows that c-command influences early antecedent retrieval, it does

not conclusively determine whether c-command information acts as a categorical constraint on antecedent retrieval or a soft constraint that only partially determines access to antecedents. The results are consistent with a retrieval mechanism that is capable of ignoring a non-c-commanding QP as a potential binder for a pronoun or one that accesses non-c-commanding QPs, but at a cost. Experiment 1 cannot distinguish between these two hypotheses because it was not designed to assess whether retrieval treats matching non-c-commanding QPs differently from NPs that do not match the pronoun. In order to answer this question Experiment 2 employs a standard interference paradigm to test whether the morphological features of a grammatically inappropriate QP interfere with antecedent retrieval.

Experiment 2: Interference

Experiment 1 indicated that antecedent retrieval distinguishes between c-commanding QPs that are eligible antecedents for a pronoun and non-c-commanding QPs that are grammatically inappropriate antecedents for that pronoun. Experiment 2 investigated whether retrieval imposes a categorical ban on non-c-commanding QPs, or whether antecedent retrieval is susceptible to partial-match interference from the morphological features of non-c-commanding QPs.

Materials

The experimental items were made up of 36 item sets, each containing 6 conditions. 4 of the 6 conditions consisted of a 2×2 factorial design and the remaining 2 conditions were lexically similar, though structurally distinct, control conditions. An example is provided in Table 6.

Test conditions

The test conditions used a 2×2 factorial design, which crossed the factors GenderMatch and Referentiality (see Table 6). All test items were transitive sentences with a plural definite subject. A singular pronoun that required an antecedent appeared as the object of the main verb. The main clause subject, which always mismatched the pronoun in number, was modified by a relative clause. The subject of the relative clause (the *potential* antecedent) was an NP whose gender and referentiality were manipulated. The position of the potential antecedent was held constant in Experiment 2 so that the potential antecedent never c-commanded the pronoun.

The factor GenderMatch manipulated whether the potential antecedent matched or mismatched the pronoun in gender features by changing the potential antecedent's head noun. Strongly gender-biased nouns (e.g. *nurse*), or inherently gendered nouns (e.g. *widow, grandfather*) were used, many of which were selected from the list of gender-biased referential nouns normed by Kennison and Trofe (2003). Pronoun gender was held constant across test items and the number of masculine and feminine pronouns was counter-balanced across items.

Table 6
Example test and control items for Experiment 2. Regions of analysis for Experiment 2b indicated by slashes.

Condition	Sentence
Referential-Match	The troop leaders that **the girl scout** had no respect/ for had scol/ded **her**/after the incident/ at scout camp./
Referential-Mismatch	The troop leaders that **the boy scout** had no respect/ for had scol/ded **her**/ after the incident/ at scout camp./
Quantificational-Match	The troop leaders that **no girl scout** had respect/ for had scol/ded **her**/ after the incident/ at scout camp./
Quantificational-Mismatch	The troop leaders that **no boy scout** had respect/ for had scol/ded **her**/ after the incident/ at scout camp./
Control-Match	The troop leaders were sure **no girl scout**/ was afraid t/hat **she**/ would be scolded/ after the incident/ at scout camp.
Control-Mismatch	The troop leaders were sure **no boy scout**/ was afraid t/hat **she**/ would be scolded/ after the incident/ at scout camp.

The factor Referentiality manipulated whether the potential antecedent was *Quantificational*, or *Referential*. In *Referential* conditions, the potential antecedent bore the definite determiner *the*. In *Quantificational* conditions the potential antecedent bore the quantifier *no*.

We predicted that a feature-matching referential NP internal to the relative clause should be readily accessed by antecedent retrieval because establishing a co-reference relation does not require c-command between an antecedent and a pronoun. Thus, we predicted undisrupted processing of the pronoun in the *Referential-Match* conditions. Following previous work, we predicted increased processing difficulty in the conditions where the potential antecedent did not match the pronoun in gender features because the pronoun lacks a suitable antecedent (e.g. Filik, Sanford, & Leuthold, 2008; Gerrig, 1986; Osterhout & Mobley, 1995). We expected clear difficulty at the pronoun in both the *Mismatch* conditions, which did not provide a matching antecedent for the pronoun. Based on the findings from Experiment 1, we also predicted difficulty in the *Quantificational-Match* condition, where the matching QP did not c-command the pronoun from inside the relative clause.[2]

Experiment 1 did not establish whether matching, non-c-commanding QPs were inaccessible, or less activated

[2] It has been noted that some relative clauses permit *functional readings*, in which QPs appear capable of binding pronouns that they do not c-command from inside the relative clause (Cooper, 1978; Rodman, 1976; Sharvit, 1999). In the sentence *The woman that every man loves is his mother* the QP *every man* can apparently bind the pronoun *his*, producing a reading equivalent to: *Every man loves a single woman who is his mother.* The readings are generally thought to be restricted to copular or specificational constructions (Heycock, 1992). Our materials did not license functional readings because they were not copular sentences. Also, our use of direct object pronouns made functional readings less likely because those readings are most natural with possessive pronouns. Finally, although Sharvit (1999) argues that functional readings can arise in non-copular sentences, those cases are restricted to the quantifiers *every*, and *each*. Our use of the negative quantifier *no* made the likelihood of a functional reading very low. This is consistent with the unacceptability of **The woman that no man wanted to marry talked with his mother.*

than c-commanding QPs, but nevertheless accessible. If matching non-c-commanding QPs are inaccessible to retrieval, gender-match between the pronoun and QP should not affect early processing of the pronoun. Difficulty at the pronoun in the *Quantificational-Match* condition should be comparable to the *Quantificational-Mismatch* condition. The pronoun should be interpreted as though it lacks a matching antecedent in both cases. On the other hand, if matching, non-c-commanding QPs are merely less activated than c-commanding QPs, we expect to observe evidence of *interference* from the gender-matching QP. In particular, we expect to observe a pattern of *facilitatory interference*, a pattern of effects previously observed in the online licensing of a number of linguistic dependencies (e.g., Parker, 2014; Pearlmutter, Garnsey, & Bock, 1999; Vasishth, Brüssow, Lewis, & Drenhaus, 2008; Wagers, Lau, & Phillips, 2009; Xiang, Dillon & Phillips, 2009). Facilitatory interference occurs when a structurally inappropriate distractor reduces the difficulty associated with processing an unlicensed linguistic element. For example, Pearlmutter et al. (1999) found that subject-verb agreement checking is subject to facilitatory interference; reading times following the ungrammatical plural verb *were* in (13) were shorter when a pre-verbal non-subject distractor matched the verb in number (*cabinets*) than when the distractor did not (*cabinet*).

(13) The key to the cabinet(s) were rusty from years of disuse.

Some authors who have replicated these findings have argued that relative facilitation arises because the distractor is mis-retrieved as a licensor for the verb due to partial overlap with the morphological cues used to retrieve a licensor (Dillon, Mishler, Slogett, & Phillips, 2013; Tanner, Nicol, & Brehm, 2014; Wagers et al., 2009). This retrieval can occur even though the distractor does not bear the correct syntactic cues (i.e. it is not marked as a *subject*, cf. Vasishth et al., 2008).

If antecedent retrieval is subject to the same facilitatory interference as other dependencies, we expect that gender-match with the QP should facilitate processing of the pronoun in Experiment 2 at the earliest possible sign of retrieval.

Control conditions

Two Control conditions were included to provide an experiment-internal measure of successful variable binding under c-command. Control items were adapted from the test items. Control sentences were three-clause sentences in which the first verb was a propositional attitude verb whose subject was a plural NP. The plural was used to ensure that it could not be an antecedent for the singular pronoun. The subject of the second clause was a QP that served as the potential antecedent for a subject pronoun in the third clause. This pronoun was c-commanded by the QP. The predicate of the most deeply embedded clause was a passivized form of the predicate in the control item's corresponding test sentence. Passives were used to avoid introducing an additional overt argument between the QP and the pronoun.

Experiment 2a: Acceptability judgment study

We gathered acceptability ratings for the test sentences from Experiment 2 in order to determine whether the c-command constraint blocks association of the RC-internal QP with the critical pronoun.

Participants

20 participants were recruited through Amazon Mechanical Turk (12 females, mean age 31.6) and paid $3.50 for their participation. Participant eligibility was determined as in Experiment 1a.

Procedure and materials

The procedure was identical to that of Experiment 1a. In the judgment study we used 30 of the 36 items from the eye-tracking study. We included 4 sentences involving the same configuration as the control conditions, but with referential embedded subjects (*Referential-Command* sentences). In order to obtain an index of the general acceptability of variable binding, we included 6 additional sentences involving variable binding in various configurations (*Good-Binding* sentences). There were 58 filler sentences, all grammatical and of comparable complexity. This resulted in the presented sentences being roughly 70% acceptable.

Analysis

Statistical analysis was conducted as in Experiment 1a with changes made to the fixed effects to reflect the different factors: Experimental fixed effects were the sum-coded factors GenderMatch and Referentiality and their interaction. Statistical analysis was carried out on *z*-scored acceptability ratings.

Results

Average raw and *z*-scored acceptability judgment ratings for Experiment 2a are given in Table 7. Within the test conditions, a main effect of Referentiality was significant ($\beta = -0.48$, s.e. = 0.09, $t = -5.630$), due to higher average acceptability ratings for *Referential* sentences over *Quantificational* sentences. A reliable effect of GenderMatch was also observed ($\beta = -0.58$, s.e. = 0.11, $t = -5.275$). *Match* conditions were rated as more acceptable than *Mismatch* conditions. The Referentiality × GenderMatch interaction was also significant ($\beta = 0.77$, s.e. = 0.17, $t = 4.660$). The *Referential-Match* condition received the highest mean acceptability rating, while all other conditions received comparatively lower acceptability scores. The *Referential-Match* condition received a higher average rating than the *Referential-Mismatch* condition ($\beta = -0.96$, s.e. = 0.14, $t = -6.780$). The *Quantificational-Match* condition had a higher average rating than the *Quantificational-Mismatch* condition (0.38 points on the seven-point scale); this numerical difference was marginally significant ($\beta = -0.20$, s.e. = 0.11, $t = -1.790$).

Table 7
Average raw and *z*-scored acceptability ratings for items in Experiment 2a. Standard errors in parentheses.

Condition	Average raw rating	Average *z*-score
Referential-Match	4.14 (.20)	−0.27 (.09)
Referential-Mismatch	2.18 (.16)	−1.23 (.08)
Quantificational-Match	2.35 (.17)	−1.13 (.07)
Quantificational-Mismatch	1.97 (.15)	−1.34 (.07)
Control-Match	3.97 (.20)	−0.38 (.09)
Control-Mismatch	2.19 (.14)	−1.20 (.06)
Filler: Referential-Command	4.66 (.23)	−0.04 (.10)
Filler: Good-Binding	5.40 (.15)	0.32 (.07)

The *Control-Match* condition also received significantly higher acceptability ratings than the *Control-Mismatch* condition ($\beta = -1.16$, s.e. = 0.17, $t = 6.970$).

Discussion

The study tested sensitivity to the grammatical generalization that referential NPs embedded in a relative clause are acceptable antecedents for a pronoun that they do not c-command, but that QPs in the same position are not acceptable antecedents.

When the critical pronoun had a feature-matching referential antecedent, participants gave test sentences relatively high ratings. However, when the sentences lacked a feature-matching antecedent for the pronoun, as they did in the *Referential-* and *Quantificational-Mismatch* conditions, participants gave test sentences low acceptability ratings. Participants also assigned low acceptability ratings to sentences in which the potential antecedent was a matching QP that failed to c-command the pronoun. These findings are consistent with the results of Experiment 1 that suggested that grammatically appropriate feature-matching phrases were more accessible as antecedents than grammatically inappropriate phrases.

Participants gave high acceptability ratings to control sentences in which a QP matched a pronoun that it c-commanded. Filler sentences that contained other acceptable cases of bound variable anaphora also received high average acceptability scores. This confirms that participants found bound variable readings available and acceptable when they were grammatically sanctioned.

The *Quantificational-Match* condition received a slightly higher acceptability rating than the *Quantificational-Mismatch* condition (a difference of 0.38 on the 7 pt scale), much smaller than the corresponding effect in the Referential conditions. This might indicate that participants occasionally considered a bound-variable relation in the absence of c-command. It could also reflect that participants were simply more confident in rejecting sentences that did not contain an NP that matched a pronoun. Regardless of this small effect, the rating study confirms that participants judge binding between the non-c-commanding QP and the pronoun to be unacceptable on the whole.

Experiment 2b: Eye-tracking while reading

Participants

32 participants from the University of Maryland community participated in the study for pay or course credit (18 females, mean age 20.8). Participants were compensated $10 for an hour of their time. All participants had normal, or corrected-to-normal vision, and were self-reported native speakers of English.

Procedure

The 36 experimental items were distributed into 6 lists in a Latin Square design with 114 additional filler sentences. The order of each list was pseudo-randomized. The testing procedure and equipment were the same as those in Experiment 1c.

Analysis

Test sentences were divided into 5 regions of interest, as indicated in Table 6. For the test conditions, the critical region contained the critical pronoun and the three characters that preceded it. As in Experiment 1c, the three characters preceding the pronoun were included in the pronoun region to account for possible parafoveal processing of the pronoun when reading the preceding context (van Gompel & Majid, 2004). The post-pronoun region included the word or two words following the pronoun, depending on the length of the head of the following prepositional phrase. The mean skipping rate across regions and conditions was 8% with a range of 0–32%. Skipping rates were highest in the pronoun region, where the mean skipping rate was 27.5%.

Reading measures reported are the same measures reported in Experiment 1c. Statistical analysis used LMEMs with the fixed effects described in Experiment 2a. All models had a maximal random effects structure. Outlier rejection procedures were the same as in Experiment 1c.

Results

Mean raw reading times by measure and region are provided in Table 8. A full report of statistical results is presented in Table 9. Fig. 4 displays effects in the critical pronoun region.

Test conditions

Pre-critical region

In the pre-critical region, reading times did not differ significantly across conditions in first-pass time, rightbound, or second-pass time measures. Participants' total reading times were numerically lower in the *Match* conditions than in the *Mismatch* conditions. This main effect of GenderMatch was marginally significant ($t = 1.847$).

Table 8
Mean raw reading times by measure and region for test conditions in Experiment 2b. Standard errors in parentheses.

		Pre-pronoun	Pronoun	Post-pronoun	Final
First-pass	*Quant-Match*	430 (19)	305 (13)	518 (23)	490 (25)
	Quant-Mismatch	470 (20)	278 (10)	512 (23)	527 (24)
	Refer-Match	450 (18)	258 (9)	480 (21)	528 (25)
	Refer-Mismatch	444 (16)	290 (11)	528 (25)	501 (23)
Right-bound	*Quant-Match*	512 (21)	344 (15)	624 (26)	750 (42)
	Quant-Mismatch	527 (22)	325 (15)	647 (25)	780 (40)
	Refer-Match	491 (18)	277 (12)	588 (22)	778 (42)
	Refer-Mismatch	496 (17)	328 (14)	638 (26)	825 (44)
Second-pass	*Quant-Match*	337 (33)	154 (17)	373 (33)	243 (33)
	Quant-Mismatch	347 (35)	180 (18)	480 (50)	276 (37)
	Refer-Match	318 (36)	128 (17)	434 (50)	262 (34)
	Refer-Mismatch	385 (34)	191 (21)	441 (38)	274 (32)
Total time	*Quant-Match*	768 (38)	441 (22)	907 (39)	754 (42)
	Quant-Mismatch	828 (41)	471 (24)	967 (44)	789 (41)
	Refer-Match	755 (40)	387 (21)	886 (43)	786 (42)
	Refer-Mismatch	819 (38)	459 (23)	978 (44)	828 (45)

Table 9
Summary of results of mixed effects models by region and measure for reading times in Experiment 2b. Results correspond to model estimates of each fixed effect's coefficient. Random intercepts were included for subjects and items, as were by-subject and by-item random slopes for all fixed effects and their interactions. Significant coefficients ($|t| > 2$) are in bold.

		Pre-pronoun		Pronoun		Post-pronoun		Final	
		Estimate (s.e.)	*t*-value	Estimate (s.e.)	*t*-value	Estimate (s.e.)	*t*-value	Estimate (s.e.)	*t*-value
First-pass	Intercept	**443.11 (22.6)**	**19.640**	**282.08 (7.5)**	**37.42**	**508.81 (26.5)**	**19.180**	**504.40 (34.3)**	**14.756**
	Refer	6.48 (18.9)	0.343	17.87 (11.9)	1.510	12.66 (27.0)	0.469	−18.97 (25.6)	−0.741
	Match	21.70 (16.6)	1.308	2.41 (15.5)	0.160	15.05 (27.2)	0.554	−1.61 (21.0)	−0.077
	Refer × Match	44.82 (38.0)	1.179	**−55.02 (21.6)**	**−2.550**	−61.46 (42.7)	−1.441	65.65 (45.7)	1.441
Right-bound	Intercept	**500.37 (26.6)**	**18.816**	**317.71 (10.8)**	**29.476**	**625.11 (30.0)**	**20.829**	**774.64 (65.0)**	**11.926**
	Refer	29.22 (19.6)	1.494	**32.27 (16.6)**	**1.942**	23.81 (24.1)	0.989	−50.37 (36.8)	−1.369
	Match	14.30 (18.1)	0.791	16.56 (19.3)	0.859	39.08 (28.3)	1.379	23.61 (44.0)	0.536
	Refer × Match	12.95 (35.4)	0.366	**−65.55 (28.0)**	**−2.345**	−35.48 (54.9)	−0.647	16.62 (99.0)	0.168
Second-pass	Intercept	**349.30 (47.9)**	**7.291**	**157.68 (20.4)**	**7.727**	**424.99 (59.2)**	**7.201**	**266.34 (48.7)**	**5.465**
	Refer	−11.74 (32.9)	−0.357	14.19 (15.2)	0.933	−12.83 (33.6)	−0.382	−12.91 (31.3)	−0.412
	Match	34.48 (29.8)	1.158	**48.92 (16.4)**	**2.975**	54.81 (31.3)	1.747	20.59 (30.3)	0.679
	Refer × Match	−41.04 (78.6)	−0.522	−27.90 (31.9)	−0.875	72.28 (76.6)	0.943	39.61 (78.2)	0.507
Total-time	Intercept	**788.77 (56.4)**	**13.979**	**438.46 (26.1)**	**16.797**	**938.44 (65.8)**	**14.255**	**781.25 (67.0)**	**11.657**
	Refer	12.32 (37.6)	0.327	32.58 (27.9)	1.168	4.73 (38.1)	0.124	−49.84 (34.8)	−1.431
	Match	62.19 (33.7)	1.847	49.08 (25.2)	1.947	**79.97 (36.7)**	**2.177**	22.17 (46.4)	0.478
	Refer × Match	6.62 (83.3)	0.079	−37.97 (55.9)	−0.679	18.08 (91.3)	0.198	29.90 (96.2)	0.311

Critical pronoun region

In first-pass times, there were no significant main effects, but RTs were characterized by a significant Referentiality × GenderMatch interaction ($t = -2.780$). Pairwise comparisons revealed that gender match with a referential NP facilitated processing of the pronoun: the *Referential-Match* condition was read more quickly than the *Referential-Mismatch* condition, ($\beta = 29.94$, s.e. = 14.55, $t = 2.06$). Gender-match with the QP did not have the same facilitatory effect. Reading times in the *Quantificational-Match* condition were higher numerically than in the *Quantificational-Mismatch* condition, but this pairwise effect was marginally significant ($t = -1.780$).

The Referentiality × GenderMatch interaction was also significant in right-bound times ($t = -2.345$). Pairwise comparison revealed a reliable effect of GenderMatch between the two *Referential* conditions, due to increased reading times in the *Referential-Mismatch* condition ($\beta = -50.77$, s.e. = 19.79, $t = 2.566$). Pairwise comparison between *Quantificational* conditions showed no reliable difference between the *Quantificational-Match* and *Quantificational-Mismatch* conditions ($|t| < 1$). In second-pass and total times, *Match* conditions were read more quickly on average than *Mismatch* conditions. This main effect was significant in second-pass measures ($t = 2.975$), but only marginally significant in total times ($t = 1.947$). GenderMatch significantly facilitated second-pass processing in the *Referential* conditions ($\beta = 62.54$, s.e. = 21.65, $t = 2.889$), but it did not have a significant facilitative effect in the *Quantificational* conditions ($\beta = 34.20$, s.e. = 21.26, $t = 1.608$). A similar pattern of effects was observed in pairwise comparison of total reading times. Total times in the *Referential-Match* condition were shorter on average than in the *Referential-Mismatch* condition ($\beta = 65.84$, s.e. = 28.64, $t = 2.299$), but the numerical trend of facilitation in the *Quantificational* conditions was not significant ($t < 1$).

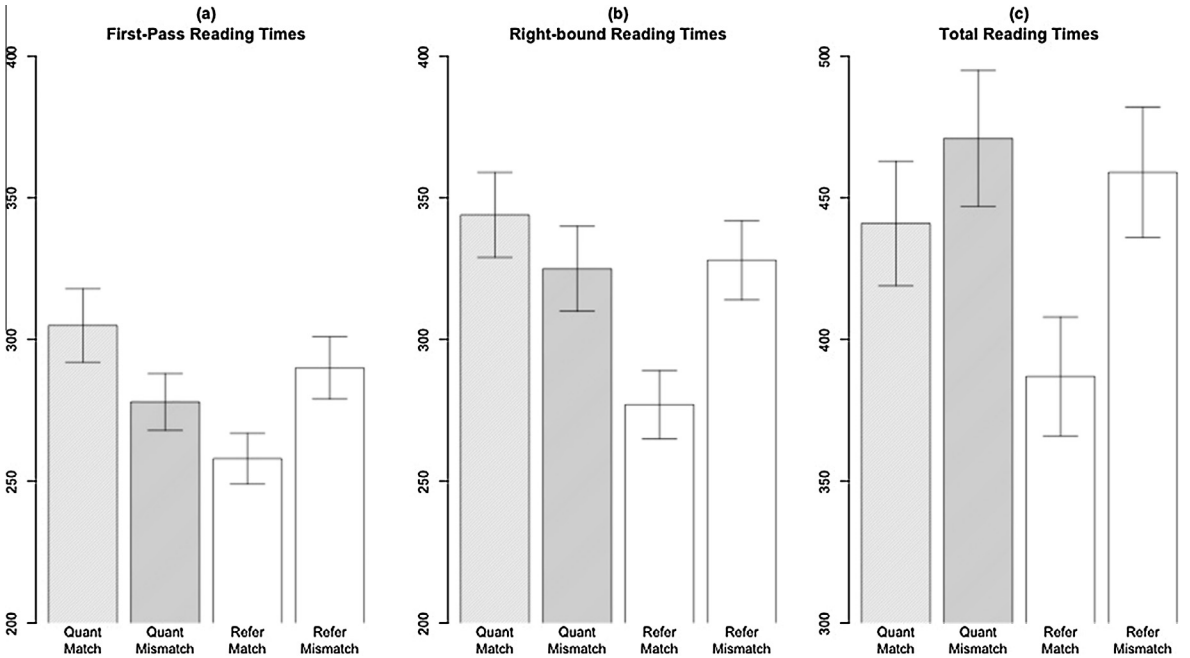

Fig. 4. (a) First-pass reading times, (b) right-bound reading times, and (c) total reading times at the pronoun region in Experiment 2b. Error bars indicate one standard error of the mean.

Post-pronoun region

The pattern of effects in early measures immediately following the pronoun was similar to the pattern observed in the pronoun region, but there were no significant effects in first-pass and right-bound reading times. In second-pass times, there was a marginally significant main effect of GenderMatch ($t = 1.747$). Visual inspection revealed that this main effect was primarily due to a trend towards an effect of GenderMatch in the pairwise comparison between the *Quantificational* conditions. We conducted pairwise comparisons even though there was not an interaction. Average reading times were shorter in the *Quantificational-Match* condition than in the *Quantificational-Mismatch* condition; this difference was marginally significant ($\beta = 90.95$, s.e. = 50.75, $t = 1.792$). The pairwise difference between the two *Referential* conditions was also not significant ($|t| < 1$). Average total reading times in the *Match* conditions were significantly shorter than those in *Mismatch* conditions ($\beta = 79.97$, s.e. = 36.74, $t = 2.177$). Average reading times were shorter in the *Quantificational-Match* condition than in the *Quantificational-Mismatch* condition; this difference was marginally significant ($\beta = 89.01$, s.e. = 52.51, $t = 1.695$). The pairwise difference between the two *Referential* conditions was also not significant ($t = 1.106$).

Sentence-final region

In the sentence-final region there were no significant effects.

Control conditions

Average raw reading times by region, condition, and measure are reported in Table 10.

Pre-critical region

In the pre-critical region the average second-pass and total reading times were numerically lower in the *Control-Match* condition than the *Control-Mismatch* condition. These numeric differences were only marginally

Table 10
Raw mean reading times by measure and region for control conditions in Experiment 2b. Standard error in parentheses.

		Pre-pronoun	Pronoun	Post-pronoun	Final
First-pass	Control-Match	394 (16)	308 (10)	484 (20)	466 (19)
	Control-Mismatch	401 (15)	307 (11)	488 (18)	482 (22)
Right-bound	Control-Match	394 (16)	332 (12)	567 (22)	551 (22)
	Control-Mismatch	399 (15)	343 (13)	604 (21)	570 (29)
Second-pass	Control-Match	338 (31)	210 (20)	413 (35)	363 (34)
	Control-Mismatch	485 (69)	277 (27)	540 (45)	392 (43)
Total time	Control-Match	725 (34)	497 (21)	892 (38)	838 (40)
	Control-Mismatch	832 (48)	555 (27)	1029 (45)	855 (41)

significant in second-pass times (β = 114.62, s.e. = 61.75, t = 1.856) and total times (β = 114.20, s.e. = 67.22, t = 1.699).

Critical pronoun region

There were no significant effects in any measure at the critical pronoun.

Post-pronoun region

In the post-pronoun region there were no significant effects in first-pass, right-bound, or second-pass times. However, participants' total reading times were reliably longer in the *Control-Mismatch* condition than in the *Control-Match* condition (β = 134.03, s.e. = 62.76, t = 2.136).

Pronoun + 2 and final regions

There were no significant effects across any measure in the final two regions.

Discussion

Experiment 2 used an interference paradigm to determine whether antecedent retrieval accesses grammatically inappropriate feature-matching QPs. The study contrasted the effect that morphological feature-match had on the accessibility of a referential NP that did not c-command the pronoun with the effect of feature-match on the accessibility of a QP in the same position. Conditions with gender mismatching NPs and QPs provided comparison conditions for measurement of retrieval failure. An additional pair of Control conditions provided an index of grammatically appropriate QP-pronoun binding under c-command.

Offline ratings in Experiment 2a confirmed that participants accept coreference between a feature-matching referential NP and a pronoun that it does not c-command, but, show little evidence of accepting a bound variable interpretation involving a non-c-commanding QP. Gender-match between the QP and the pronoun led to a reliable, but small increase in the acceptability of such sentences.

The eye-tracking study tested whether the contrast between referential and quantificational NPs observed in acceptability ratings would have a similar effect in early reading time measures. In Referential conditions gender match between the pronoun and antecedent facilitated processing at the pronoun. In Quantificational conditions, gender match with the pronoun did not facilitate processing at the pronoun in these measures. This pattern of effects shows that upon encountering a pronoun, comprehenders can rapidly access a feature-matching referential NP that does not c-command the pronoun as a potential antecedent. Comprehenders do not access a QP in an identical position, irrespective of morphological feature-match. These findings suggest that relational information has an immediate effect on the automatic retrieval process initiated when the parser first encounters a pronoun.

The contrast seen in first-pass and right-bound measures persisted into second-pass and total time measures at the pronoun, but was more equivocal. Gender match continued to facilitate processing between Referential conditions, but it also appeared to have a weak facilitative effect in Quantificational conditions. The presence of a main effect of gender match in *Referential* and *Quantificational* conditions alike suggested that Referential and Quantificational conditions might have been processed similarly, but pairwise comparisons do not bear this out. Facilitatory effects at the pronoun were consistently reliable across measures in the Referential conditions. Facilitatory effects at the pronoun only arose in second-pass and total time measures in Quantificational conditions and these effects were not reliable. A marginally significant facilitatory effect of gender match was also observed in Quantificational conditions in the post-pronoun region. Referential conditions did not show reliable facilitation in this region. This effect provides further confirmation that the gender match effects observed in Quantificational conditions is different from the gender match effect in Referential conditions.

Immediate facilitation in the *Referential-Match* condition is expected and is consistent with successful retrieval of the referential NP and resolution of a coreference relation with the pronoun. We reason that the delayed effect of gender-match in the *Quantificational* conditions does not reflect interference during initial antecedent retrieval because the facilitation did not occur in first-pass or right-bound times at either the pronoun or post-pronoun region, whereas gender-match had a clear effect in the *Referential* conditions.

The presence of the matching QP did not impact initial retrieval, but the delayed effect of gender match indicates that the matching QP may have influenced some stage of processing. We suggest that this influence occurs after retrieval fails to find a suitable antecedent for the pronoun in Quantificational conditions and the parser must find a way to accommodate an unlicensed pronoun. The parser could use the matching non-c-commanding QP during later processing in two ways. It could use the QP as part of the process of accommodating a referential interpretation of the pronoun that is disjoint from the QP or it could attempt to establish a binding relation with the QP in violation of the c-command constraint.

Under the first option, the parser would assume that antecedent-less pronoun must refer to a new individual that had not been appropriately introduced in the discourse context (Filik, Sanford, & Leuthold, 2008; Nieuwland, 2014). The process of accommodating the coercion of the new referent would involve making a *bridging inference* to update the discourse representation (Garrod & Sanford, 1981; Greene et al., 1994; Haviland & Clark, 1974; McKoon, Gerrig, & Greene, 1996). The presence of a feature-matching QP could ease the process of accommodation by providing salient and relevant properties that can be predicated of the new extra-sentential referent (e.g., Gerrig & O'Brien, 2005). Importantly, although a matching QP can facilitate processing of the pronoun under this account, the QP is never considered as an antecedent for the pronoun itself.

Under the second option pronoun resolution might occasionally permit the comprehender to consider the matching QP as a binder after antecedent retrieval failed to find a grammatical antecedent. This option is consistent with the possibility that the increased selection of

quantificational paraphrases in the *Quantificational-But* condition in Experiment 1b might reflect consideration of a QP-pronoun binding relation. We point out, however, that if comprehenders do consider non-c-commanding matching QPs as antecedents for a pronoun, it appears that they do not do so reliably.

When a QP c-commanded the critical pronoun in the *Control* conditions gender match facilitated processing. This facilitation was observed in second-pass and total time measures immediately following the pronoun, but not in earlier measures. The differential time-course of mismatch effects between test and control conditions raises questions regarding the accessibility of quantificational NPs generally. We do not have an account of why the effects of gender match emerged later in the Control conditions than in the Referential conditions. We speculate that the delay may relate to differences between processing subject and object pronouns, or the different constructions used.

One possible interpretation of this delay is that it reflects a general delay in, or dis-preference for, attempting bound variable readings. On this interpretation the delayed effects in the test conditions might also reflect delayed initial retrieval of QPs. We consider this option unlikely for two reasons. First, prior studies (Carminati et al., 2002; Cunnings et al., 2014) and Experiment 1c of this paper provide evidence that variable binding displays a time-course comparable to co-reference. Second, we do not think it is appropriate to compare the time course of the gender match effects in the Quantificational test conditions with those in the Control conditions because these conditions were not closely matched. It seems more likely that whatever the explanation for the delayed gender match in the Control conditions is independent of the delayed effects of gender match in the Quantificational test conditions.

General discussion

Summary of results

The present study assessed whether antecedent retrieval accesses non-c-commanding feature-matching QPs in violation of the c-command constraint on quantificational binding (Reinhart, 1983). This overarching question was divided into two sub-questions of interest. First, we investigated whether c-command had any observable effect on the retrieval of a QP under conditions where the grammar only licenses an anaphoric dependency with a c-commanding QP. Second, we asked whether c-command acted as a categorical constraint on initial antecedent retrieval. The findings probed the predictions of cue-based models of memory access in which relational information such as c-command is difficult to encode and use as a cue for retrieval. They also serve to enrich our general understanding of which grammatical constraints are and are not obeyed in real-time sentence comprehension.

Offline judgment studies corroborated the generalization from the linguistics literature that c-command constrains the binding relation between a QP and a pronoun. In Experiments 1a and 2a, sentences containing a pronoun and a gender-matching but non-c-commanding NP were rated as less acceptable when the NP was quantificational than when it was referential. In most cases, the sentences with non-c-commanding QPs were judged as equally unacceptable as sentences that contained no feature-matching NP at all. In Experiment 1b the majority of participants did not entertain a binding dependency between a pronoun and a non-c-commanding QP, but readily did so with a c-commanding QP. A small subset of participants appeared to occasionally consider readings that were inconsistent with the c-command constraint, but it was unclear whether this was due to demands of the forced-choice paraphrase task, or whether these participants occasionally entertained bound readings of pronouns in the absence of c-command.

Experiment 1c showed that participants experienced more difficulty processing a pronoun when the only feature-matching phrase in the sentence was a non-c-commanding QP than when the matching phrase was a c-commanding QP or a referential NP. Experiment 2b demonstrated that gender match between a pronoun and a non-c-commanding QP did not facilitate the early stages of pronoun resolution. We concluded from the results of Experiment 2b that antecedent retrieval does not access structurally inappropriate QPs. Although initial retrieval did not exhibit sensitivity to interference, it did appear that later processing may be influenced by a structurally inappropriate QP.

The role of relations in retrieval

The results suggest that antecedent retrieval can use relational information to distinguish QPs that could license bound-variable pronoun readings from QPs that cannot. The results pose a *prima facie* challenge for cue-based retrieval mechanisms. If bound-variable interpretations depend on a relation such as c-command and if cue-based mechanisms are incapable of using relations as cues, it would appear that such models are unequipped to explain the retrieval sensitivity observed here. There are two options that could be pursued to accommodate these results. First, we could reject the assumption that a direct-access cue-based mechanism is used to retrieve antecedents for pronouns. Alternatively, we could attempt to encode a proxy feature for the constraint that does not require making reference to relational information during retrieval. We explore these two possibilities below.

The experiments above could be interpreted as providing evidence for a retrieval method other than direct access. One possible retrieval mechanism is a serial search procedure (McElree & Dosher, 1989, 1993; McElree et al., 2003; Sternberg, 1966). Serial search procedures operate over a specified domain and iteratively consider individual items in that domain according to some ordering function. A search-driven model of antecedent retrieval could be an attractive alternative to direct access if it could systematically exclude non-c-commanding QPs from the set of NPs searched.

Some extant theories of anaphor resolution, such as Discourse Prominence Theory (Gordon & Hendrick, 1997, 1998b), hold that retrieval iteratively searches through a list of NPs that are ranked according to their *prominence*,

where prominence is calculated using linear order and level of syntactic embedding. Current prominence-driven search procedures do not provide a straightforward account of our findings, because prominence does not distinguish inaccessible QPs from accessible referential NPs if they occupy the same non-c-commanding position.

An alternative search-driven model of antecedent retrieval is one that traverses a path of c-commanding positions starting from the pronoun. Some authors have advocated the use of such a mechanism for other structurally restricted dependencies (e.g., Dillon, 2014 for reflexive licensing). This serial search procedure permits a straightforward implementation of a c-command constraint, but the same mechanism is not well suited for modeling the retrieval of referential antecedents for pronouns, because referential antecedents are not subject to a c-command constraint. Yet Experiments 1 and 2 demonstrate that comprehenders show no difficulty in accessing non-c-commanding referential antecedents. If antecedent retrieval for c-commanding antecedents employs a serial search strategy, then a supplemental retrieval procedure would be needed to identify non-c-commanding referential antecedents. A 'dual retrieval' strategy of this sort could deploy two serial search procedures in parallel: one procedure that traversed the c-command path to identify potential binders and a second procedure that searched the linear string or a list of NPs ranked according to their prominence for potential referential antecedents (Clark & Sengul, 1979; Ehrlich & Rayner, 1983; Gordon & Hendrick, 1998b; O'Brien, 1987).

Meanwhile, accommodating the results within a cue-based architecture would require a reconceptualization of the constraint on bound variable interpretations so that it did not require retrieval mechanisms to use relational information such as c-command. This would entail recoding the relational constraint into a constraint on feature-match. We do not endorse prominence-driven models as the solution to explaining the inaccessibility of non-c-commanding QPs, but we note that the calculation of prominence incorporates an insight that may be important for accounting for the observed pattern of effects. This is the idea that the accessibility of referents can change dynamically over the course of an incremental parse, in response to syntactic triggers.

Identifying potential binders of a pronoun requires calculation of the relative position of a QP and the pronoun only if evaluation of the QP's eligibility occurs at retrieval time. We have thus far assumed that this approach is necessary for pronouns, since they cannot in general be anticipated before they appear in the input. However, below we explore the possibility that it might be possible over the course of incremental parsing to exploit information in the linguistic context to pre-emptively mark a QP as inert for all subsequent retrievals. This proposal recasts the accessibility of a QP in terms of the state of the incremental parse, rather than the relation it bears to a particular item in the syntactic representation.

The proposal begins from two observations: First, a quantifier takes *scope*, i.e. it has the ability to bind a pronoun, so long as it c-commands the portion of the phrase marker that is currently under construction by the incremental parser. Second, a quantifier can lose the ability to take scope as the incremental construction of a sentence or discourse progresses (see, e.g., Gordon, Grosz, & Gilliom, 1993; Gordon & Hendrick, 1998a, 1998b; Heim, 1982, 1983; Kamp & Reyle, 1993). A quantifier that was eligible to bind earlier pronouns loses its ability to bind subsequent pronouns as soon as the parser begins building material outside of the QP's c-command domain. The right edge of a QP's scope domain can, under most circumstances, be defined structurally as the end of the right branch that includes the QP's rightmost sister. In (14), the right brackets delimit the respective edges of the QPs' c-command domains. The domain of *any janitor* extends to the end of the first conjunct, while the domain of *no girl scout* extends to the end of the RC (immediately after *for*).

(14a) [Kathi didn't think that any janitor liked his job] *but*...

(14b) The troop leaders that [no girl scout had any respect for] *scolded*...

An incremental parser could potentially recognize the edge upon encountering the connective *but* or the verb *scolded* because these words must be attached at a higher level of the structure. Having identified the edge of the scope domain, the parser would have enough information to know that the QP could no longer license bound variable dependencies with subsequent material. An optimal memory encoding and retrieval system could and should at this point render the item ineligible for later retrievals associated with dependency creation.

We suggest that an item's status as a viable antecedent for anaphoric relations could be encoded as a feature, ACCESSIBLE, on a chunk in memory. The ACCESSIBLE feature could be used as a retrieval cue: only those NPs that match the feature ACCESSIBLE should be retrieved as potential antecedents for a pronoun. Under this proposal, referential NPs should always bear the ACCESSIBLE feature; they never lose their ability to serve as antecedents for later pronouns. QPs, on the other hand, bear an ACCESSIBLE feature that lasts only as long as the parser is elaborating their scope domain. A QP would be introduced into the representation of a sentence bearing the ACCESSIBLE feature, but the parser should delete the feature once it reaches the edge of the QP's scope domain.[3] We propose that this can be achieved with an automatic, dynamic update procedure: whenever the parser shifts to a higher level of embedding from its previous position, it should retrieve all QPs at the last level and de-activate their ACCESSIBLE features. Retrieval of this subset of relevant QPs could be achieved if QPs bore features that tracked the branch of the syntactic tree to which they were attached (see Kush, 2013 for a more detailed description of this proposal).

One advantage of this account is that, unlike the dual-search procedure sketched above, it does not require distinct retrieval mechanisms for referential and

[3] The distribution of the ACCESSIBLE feature very closely tracks whether the QP that bears the feature would have a corresponding *discourse referent* in a *subuniverse* of a Discourse Representation Theory model (see Kamp & Reyle, 1993).

quantificational antecedents. A single retrieval step identifies all ACCESSIBLE noun phrases, referential and quantificational alike, that could serve as grammatically sanctioned antecedents for a pronoun. We elaborate on the implications of this below.

The account of retrieval in terms of the feature ACCESSIBLE provides a method to block access to non c-commanding QPs that is compatible with a cue-based architecture, because it does not require encoding of relational or item-to-item information on individual items. A QP's accessibility is not evaluated in relation to any individual item, but rather in relation to the state of the left-to-right parser. The account does not offer a feature-based translation of c-command *per se*, but c-command and syntactic structure nevertheless play an integral role in determining feature assignment. Reference to c-command resides in the update function for the ACCESSIBLE feature.

The proposal above achieves sensitivity to a c-command constraint, but it does not provide a complete account of the range of distributional constraints on potential antecedents for pronouns as determined by c-command. We discuss two remaining considerations.

ACCESSIBILITY is conceived of as a simple precondition on antecedent-pronoun relations. It marks whether a phrase could, in principle, support some kind of anaphoric relation with an unspecified anaphor in the subsequent discourse. It cannot be used to enforce additional syntactic constraints, such as anti-locality constraints like Principle B (Chomsky, 1981). For example, *no man* would be ACCESSIBLE at the processing of *him* in (15), but binding is blocked in this configuration.

(15) No man$_{[\text{ACCESSIBLE}]}$ hurt him.

Recent work has suggested that pronoun resolution is not susceptible to interference from Principle B-violating NPs (Nicol, 1988; Nicol & Swinney, 1989; Chow et al., 2014; Lee & Williams, 2008, though see Badecker & Straub, 2002), which suggests that there must be a separate implementation of the constraint that operates successfully online.

The second issue relates to the fact that retrieval does not distinguish c-commanding referential NPs from non-c-commanding referential NPs; both types are always marked as ACCESSIBLE. This might be problematic because, although c-command is not required for a referential NP to co-refer with a pronoun, c-command is important for determining whether referential NPs can bind a pronoun.

The pronoun *him* in (16) can be coreferential with *the science teacher* if both the pronoun and the NP point to the same entity in the discourse context through the use of a referential index, as depicted in (16a). The pronoun could also be bound by the main subject, as depicted in (16b); on this reading, its interpretation would co-vary with the interpretation of any NP inserted into the main subject position. Because *the science teacher* occupies this position, it binds the pronoun.

(16) The science teacher thought that the kids respected him.
  a. Coreference:
     The science teacher$_1$ thought that the kids respected him$_1$.
  b. Binding:
     x = the science teacher: [x thought that the kids respected x].

It can be difficult to distinguish between readings in which a referential NP corefers with or binds a pronoun, but the effects of binding can be seen in readings of pronouns under ellipsis (e.g. Ross, 1967; Sag, 1976). In (17), the bracketed constituent *thought that the kids respected him* serves as the antecedent for ellipsis of the VP in the second conjunct (represented as an empty constituent [$_{\text{VP}}$ _]). Because ellipsis is subject to a parallelism condition (Fiengo & May, 1994; Hestvik, 1995), the pronoun in the elided constituent must be interpreted in the same manner as the pronoun in the VP in the first conjunct. If the pronoun is referential in the first conjunct, as in (17a), the elided pronoun must bear the same index and refer to the same individual (*the science teacher*). On this reading, the music teacher believes that the kids respect the science teacher. If the pronoun is bound by the local subject in the first conjunct (as in 17b), the elided pronoun must be bound by the local subject in the second conjunct (*the music teacher*). Under the bound reading, the music teacher believes himself to be respected.

(17) The science teacher <thought that the kids respected him>, and the music teacher did [$_{\text{VP}}$ ____], too.
  a. Referential Reading:
     The science teacher$_1$ thought that the kids respected him$_1$ &
     The music teacher$_2$ thought that the kids respected him$_1$
  b. Bound Reading:
     x = the science teacher: [x thought that the kids respected x] &
     y = the music teacher: [y thought that the kids respected y],

As with QP-pronoun binding, a referential NP can only bind a pronoun that it c-commands. In (18) *the science teacher* no longer c-commands *him* in the first VP, so it cannot bind the pronoun. As a result, it cannot license a bound reading in the elided VP. The second conjunct can only mean that the music teacher thinks that the kids respect the science teacher.

(18) The parents [that met the science teacher] <thought that the kids respected him>, and the music teacher did [$_{\text{VP}}$ ____], too.

These facts indicate that a comprehensive model of pronominal resolution must provide a mechanism for computing NP-pronoun binding dependencies. If a referential NP's c-command relation with a pronoun is not visible to retrieval, the retrieval mechanism cannot determine whether that NP is a potential binder for a pronoun. This entails that the parser would need to check post-retrieval whether the retrieved NP c-commands the pronoun in order to ascertain whether binding is possible. We note

that such a checking process is consistent with the assumption of many models of pronominal resolution that incorporate a post-retrieval stage where interpretive decisions are made and evaluated (e.g., Sanford & Garrod, 1998).

Cue-combinatorics

Many models of cue-based retrieval assume that retrieval cues combine equally and simultaneously to activate potential targets (e.g., Clark & Gronlund, 1996; Lewis & Vasishth, 2005). One prediction of models that incorporate a linear cue-combinatorics scheme is that retrieval should be susceptible to partial-match interference. If antecedent retrieval uses both gender information and the ACCESSIBLE feature as equally weighted cues, we should expect to observe partial match interference effects not found in our studies. The absence of partial-match interference from structurally inaccessible distractors suggests that the ACCESSIBLE feature performs some kind of gating function. As discussed by Van Dyke and McElree (2011), this gating effect could be modeled by preferentially-weighting the ACCESSIBLE feature over morphological features in a linear cue combinatorics scheme, or by using a multiplicative combinatorial scheme. Under a weighted linear scheme, morphological match would still influence activation of structurally inappropriate distractors, but such effects would be negligible relative to structural cue match. Alternatively, a multiplicative scheme would impose the strong constraint that the activation of any NP that mismatched the probe on structural cues would be null. At present we do not have the power to discern between these two options.

Conclusion

Our studies asked whether antecedent retrieval respects the c-command constraint on licensing bound-variable pronouns. Results from two eye-tracking studies suggest that retrieval not only distinguishes structurally inappropriate from structurally appropriate QPs, but that c-command information could serve as a categorical filter on the initial retrieval of quantified phrases as potential antecedents. We found weak evidence in offline and late eye-tracking measures that the gender of a non-c-commanding QP could exert a small influence on the processing of a pronoun. We attributed these effects to later interpretive processes that occur after the failure of initial antecedent retrieval.

The configurational nature of the c-command constraint on bound-variable pronoun licensing presents a difficult, though not insuperable, challenge to modeling antecedent retrieval using cue-based models of retrieval. The current results attest to the importance of this relation as a real-time constraint on anaphor resolution. Consequently, it appears that the language processing architecture must have a means for accommodating the distinctions that this relational constraint encodes. We hypothesized that a dynamically updated cue that tracks an NP's accessibility could be used to achieve grammatical sensitivity. This account has the advantage of avoiding the need to encode the c-command constraint in the vocabulary of retrieval cues.

Acknowledgments

Earlier versions of this work were presented at the CUNY Human Sentence Processing Conference in 2009 and 2012. We thank Pedro Alcocer, Brian Dillon, Clint Johns, Lyn Frazier, Norbert Hornstein, Ellen Lau, Julie Van Dyke, Shravan Vasishth, and Alexander Williams for helpful discussion. This work was supported in part by NSF grant BCS-0848554 to Colin Phillips and NSF IGERT DGE-0801465 to the University of Maryland.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
Anderssen, J. (2011). *Quantification, misc*. Doctoral dissertation. Amherst: University of Massachusetts.
Avrutin, S. (1994). *Psycholinguistic investigations in the theory of reference*. Doctoral dissertation. Massachusetts Institute of Technology.
Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.
Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 748–769.
Barker, C. (2012). Quantificational binding does not require c-command. *Linguistic Inquiry, 43*, 614–633.
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*, 255–278.
Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of 'each' and 'every'. In A. Szabolcsi (Ed.), *Ways of scope taking* (pp. 71–107). Dordrecht: Kluwer.
Bosch, P. (1983). *Agreement and anaphora: A study of the role of pronouns in syntax and discourse*. London: Academic Press.
Büring, D. (2005). *Binding theory*. Cambridge, UK: Cambridge University Press.
Carminati, M. N., Frazier, L., & Rayner, K. (2002). Bound variables and c-command. *Journal of Semantics, 19*, 1–34.
Chomsky, N. (1981). *Lectures on government and binding*. Berlin: Mouton de Gruyter.
Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology, 5*, 630.
Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*, 37–60.
Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition, 7*, 35–41.
Cooper, R. (1978). Variable binding and relative clauses. In *Formal semantics and pragmatics for natural languages* (pp. 131–169). Netherlands: Springer.
Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition, 11*, 283–294.
Cunnings, I., Patterson, C., & Felser, C. (2014). Variable binding and coreference in sentence comprehension: Evidence from eye movements. *Journal of Memory and Language, 71*, 39–56.
Dillon, B. (2014). Syntactic memory in the comprehension of reflexive dependencies: An overview. *Language and Linguistics Compass, 8*, 171–187.
Dillon, B., Mishler, A., Slogett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language, 69*, 85–103.
Drummond, A. (2011). *IbexFarm (Version 0.3) [Software]*. <http://spellout.net/ibexfarm/>.
Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior, 22*, 75–87.
Fiengo, R. (1977). On trace theory. *Linguistic Inquiry, 8*, 35–62.

Fiengo, R., & May, R. (1994). *Indices and identity*. Cambridge, MA: MIT Press.
Filik, R., Sanford, A. J., & Leuthold, H. (2008). Processing pronouns without antecedents: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 20*, 1315–1326.
Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Availability versus accessibility. *Journal of Memory and Language, 56*, 357–383.
Frazier, L., & Clifton, C. Jr, (2000). On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research, 29*, 125–140.
Garnham, A., Oakhill, J., Ehrlich, M. F., & Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory and Language, 34*, 41–62.
Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language, 33*, 39–68.
Garrod, S., & Sanford, A. J. (1981). Bridging inferences and the extended domain of reference. *Attention and Performance IX*, 331–346.
Garrod, S. C., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language, 42*, 526–544.
Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27*, 699–717.
Gerrig, R. J. (1986). Process models and pragmatics. *Advances in Cognitive Science, 1*, 23–42.
Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes, 39*, 225–242.
Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–65.
Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 17*, 311–347.
Gordon, P. C., & Hendrick, R. (1997). Intuitive knowledge of linguistic co-reference. *Cognition, 62*, 325–370.
Gordon, P. C., & Hendrick, R. (1998a). Dimensions of grammatical coreference. In *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 424–429).
Gordon, P. C., & Hendrick, R. (1998b). The representation and processing of coreference in discourse. *Cognitive Science, 22*, 389–424.
Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language, 33*, 511–526.
Guo, F., Foley, C., Chien, Y.-C., Chiang, C.-P., & Lust, B. (1996). Operator-variable binding in the initial state: A cross-linguistic study of VP-ellipsis structures in Chinese and English. *Cahiers de Linguistique-Asie Orientale, 25*, 334.
Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*, 512–521.
Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Doctoral dissertation. Amherst: University of Massachusetts.
Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use and interpretation of language* (pp. 164–189). Berlin: De Gruyter.
Hestvik, A. (1995). Reflexives and ellipsis. *Natural Language Semantics, 3*, 211–237.
Heycock, C. (1992). Layers of predication and the syntax of the copula. *Belgian Journal of Linguistics, 7*, 95–123.
Kamp, H., & Reyle, U. (1993). *From discourse to logic*. Dordrecht: Kluwer.
Kayne, R. (1994). *The anti-symmetry of syntax*. Cambridge, MA: MIT Press.
Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research, 32*, 355–378.
Kush, D. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing*. Doctoral dissertation. College Park: University of Maryland.
Ladusaw, William. (1980). *Polarity sensitivity as inherent scope relations*. New York & London: Garland Publishing.
Lee, M.-W., & Williams, J. N. (2008). *The Role of grammatical constraints in intra-sentential pronoun resolution*. London: London Metropolitan University/Cambridge University manuscript.
Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–419.
Lewis, R., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science, 10*, 447–454.
May, R. (1977). *The grammar of quantification*. Doctoral dissertation. Massachusetts Institute of Technology.
McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*, 111–123.
McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation, 46*, 155.
McElree, B., & Dosher, B. (1989). Serial position and set size in short-term memory: Time course of recognition. *Journal of Experimental Psychology: General, 18*, 346–373.
McElree, B., & Dosher, B. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122*, 291–315.
McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*, 67–91.
McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: Some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 919.
Nicol, J. (1988). *Coreference processing during sentence comprehension*. Doctoral dissertation. Massachusetts Institute of Technology.
Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*, 5–19.
Nieuwland, M. S. (2014). Who's he?: Event-related brain potentials and unbound pronouns. *Journal of Memory and Language, 76*, 1–28.
O'Brien, E. J. (1987). Antecedent search processes and the structure of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 278.
Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739–773.
Öztekin, I., & McElree, B. (2010). Relationship between measures of working memory capacity and the time course of short-term memory retrieval and interference resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 383.
Parker, D. (2014). *The cognitive basis for encoding and navigating linguistic structure*. Doctoral dissertation. College Park: University of Maryland.
Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*, 427–456.
Poesio, M., & Zucchi, A. (1992). On telescoping. In C. Barker & D. Dowty (Eds.), *Proceedings of the second conference on semantics and linguistic theory II* (pp. 347–366). Columbus, Ohio: Ohio State University. Working Papers in Linguistics.
Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.
Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.
Roberts, C. (1989). Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy, 12*, 683–721.
Rodman, R. (1976). Scope phenomena, 'movement transformations', and relative clauses. In B. Partee (Ed.), *Montague grammar*. New York: Academic Press.
Ross, J. R. (1967). *Constraints on variables in syntax*. Doctoral dissertation. Massachusetts Institute of Technology.
Sag, I. A. (1976). *Deletion and logical form*. Doctoral dissertation. Massachusetts Institute of Technology.
Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes, 26*, 159–190.
Sanford, A. J., & Garrod, S. C. (2005). Memory-based processing and beyond. *Discourse Processes, 39*, 205–224.
Schütze, C. T., & Sprouse, J. (2014). Judgment data. In R. Podesva & D. Sharma (Eds.), *Research methods in linguistics*. Cambridge: Cambridge University Press.
Shapiro, L., & Hestvik, A. (1995). On-line comprehension of VP-Ellipsis: Syntactic reconstruction and semantic influence. *Journal of Psycholinguistic Research, 24*, 517–532.
Sharvit, Y. (1999). Functional relative clauses. *Linguistics and Philosophy, 22*, 447–478.
Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods, 43*, 155–167.
Sternberg, S. (1966). High speed scanning in human memory. *Science, 153*, 652–654.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language, 76*, 195–215.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*.

van Gompel, R. P., & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition, 90*, 255–264.

Vasishth, S., Brüssow, S., Lewis, R., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science, 32*, 685–712.

Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*, 206–237.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language, 108*, 40–55.