Configuration-sensitive Retrieval: Resisting Interference in Processing Bound Variable Pronouns

Dave Kush*
Haskins Laboratories, New Haven, CT 06511, USA

Jeffrey Lidz
Colin Phillips
Department of Linguistics, University of Maryland, College Park, MD 20742, USA

*Corresponding author:
Dave Kush
Haskins Laboratories
300 George St.
New Haven, CT 06511
kush@haskins.yale.edu
Abstract

Formal theories of grammatical knowledge make extensive use of syntactic relations (e.g. c-command, Reinhart, 1983) in the description of constraints on dependency formation such as antecedent-anaphor relations. Recent research has motivated a model of processing that exploits a cue-based retrieval process in content-addressable memory (e.g. Lewis, Vasisht & 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. As a test of memory access’s ability to use relational information we investigated antecedent retrieval associated with bound variable anaphora, a form of anaphoric dependency that imposes a c-command restriction on antecedent-pronoun relations. In order to bind a pronoun, a Quantificational NP (QP, e.g., no janitor) must c-command that pronoun (e.g. Reinhart, 1983). We contrasted retrieval of QPs with the retrieval of referential NPs (e.g. the janitor) for establishing coreference, which is not subject to the same c-command constraint. In three off-line judgment studies and two eye-tracking studies, we show that non-c-commanding QPs are reliably distinguished from grammatical antecedents and that non-c-commanding QPs do not interfere with antecedent retrieval for a feature-matching pronoun. C-commanding QPs are easily accessed by retrieval, as are referential NPs regardless of their c-command relation to 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 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.

Keywords:
anaphora; binding; c-command; memory retrieval; sentence processing
Pronouns typically depend for their interpretation on previously seen items in the linguistic and non-linguistic context. In other words, pronouns require retrieval of antecedents from memory. The relations between pronouns and their antecedents are also subject to numerous constraints, which have been extensively studied in linguistics and psycholinguistics. Therefore, pronoun resolution provides a valuable test case for investigating 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 configurational constraint on antecedents that is not easily captured in otherwise well-motivated cue-based models of memory access.

Pronouns can be interpreted as either referential (1) or as bound variables (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.

In (1) the pronoun is co-referential with its antecedent the cyclist; both expressions are thought to point to the same single individual in a discourse model (see, e.g., Büring, 2005). In (2a,b) the pronouns are said to be 'bound' by their antecedents. The pronouns do not refer to a single individual in the discourse model, but rather co-vary in interpretation with the quantifier, which provides instructions on how to iterate through individuals in the discourse model.

Regardless of their interpretation, all pronouns are subject to semantic/morphological constraints on suitable antecedents (e.g., they must match their antecedents in features such as number and gender). They are also subject to constraints on the locality of their antecedents (e.g. a pronoun
cannot have a structurally prominent clause-mate antecedent, Chomsky, 1981). But whereas referential pronouns are relatively liberal in where they can find their antecedents, allowing intra- or extra-sentential linguistic antecedents, or non-linguistic antecedents, bound variable pronouns are subject to more stringent structural constraints.

In order to be bound a pronoun must stand in a particular structural configuration with respect to its antecedent. Since Reinhart (1979/1983), the relevant configurational relation is referred to as c-command (though see Barker, 2012 for a recent critique of this formulation). An item $X$ c-commands another, $Y$, 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 VP that is the QP’s sister. The pronouns cannot be bound by the feature-matching QPs in (3a,b), however, because those QPs are too deeply embedded to c-command the pronoun. Meanwhile, co-reference between the pronoun and the QP is not possible because QPs are inherently non-referential. Thus, the pronouns in (3a,b) lack a sentence-internal antecedent.

(3a) *The photographers [that no cyclist posed for] still had pictures of him.
(3b) * The photographers [who had spoken with every cyclist] took pictures of him.

The c-command constraint on bound variable pronouns is particularly interesting for models of memory encoding and access, because its standard formulation presents a potential challenge for otherwise well-motivated models of memory access. Popular cue-based models assume that memory access makes exclusive use information that can be encoded as an inherent feature on an item as retrieval cues (e.g. Gillund & Shiffrin, 1984; Anderson, 1990; Lewis, Vasishth & Van
Dyke, 2006). This type of item information is often assumed to include lexical category, morphological features (e.g., number and gender), syntactic information (e.g., grammatical role), or semantic information (e.g., lexical features).

Certain constraints on antecedents are easily implemented in this kind of framework. For example, the constraint on morphological feature-match can be straightforwardly enforced because morphological features are item information. C-command relations are different in kind from morphological features. Syntactic relations are item-to-item information; they are necessarily stated over pairs of nodes in a structure. Relational information cannot easily be used in a system that is restricted to making reference to item information. Simply put: under current models, c-command relations cannot be used as cues for direct-access retrieval. In light of the centrality of c-command in the characterization of diverse linguistic constraints (e.g., Fiengo, 1977; Chomsky, 1981; Kayne, 1994), we therefore encounter a clear tension between well-motivated linguistic notions and well-motivated memory mechanisms.

Identifying c-commanding antecedents in retrieval

Under any account of pronoun interpretation, two classes of mechanisms can be distinguished: (i) mechanisms that identify potential antecedents, and (ii) mechanisms that determine the most suitable antecedent from the initial candidate set (e.g., Garrod and Sanford, 1998). We assume that the mechanism that identifies potential antecedents is the retrieval mechanism. The first question we ask is whether the antecedent retrieval mechanism can distinguish grammatically appropriate (c-commanding) potential antecedents from grammatically inappropriate (non-c-commanding) noun phrases. A simplistic cue-based retrieval model would predict that retrieval
should be incapable of making such a distinction. If retrieval were capable of distinguishing c-commanding from non-c-commanding noun phrases it would either constitute counter-evidence to a cue-based model of antecedent retrieval or it would suggest that the parser uses a feature-based proxy for c-command as a cue for retrieval (see Kush, 2013 for extensive discussion on the use of feature-based proxies for c-command constraints).

**Cue-combinatorics**

In addition to the role of c-command in determining antecedent accessibility, we address a second, related issue in this paper - the way in which information is combined to activate targets of retrieval. Many cue-based retrieval models (e.g. Lewis & Vasishth, 2005) adopt a view that the probability of retrieving an item, $p(R_i)$, is proportional to the sum of the weighted strength of association $(w_qS_{iq})$ between the item $(i)$ and a series of individual retrieval cues $(q)$. In such models it is often assumed that cue weights are equal, reflecting the assumption that no information takes priority in retrieval.

\begin{equation}
    p(R_i) = \sum w_qS_{iq}
\end{equation}

Such cue-combinatoric equations provide a rationale for frequently observed *partial-match interference* effects. A number of authors have argued that grammatically inappropriate items whose cues partially-match a retrieval probe’s cue set exert an influence on the retrieval process (see, e.g., Badecker & Straub, 2002; Lewis & Vasishth, 2005; Lewis, Vasishth & Van Dyke, 2006; Van Dyke, 2007; Van Dyke & McElree, 2011; Vasishth, Brüssow, Drenhaus & Lewis, 2008; Wagers, Lau & Phillips, 2009). Such cue-combinatorics schemes allow partial match
interference, because the contribution that each individual cue makes to the probability of retrieving an item is independent from the contribution of other cues.

It has recently been argued that a cue-combinatorics scheme that allows certain cues to 'gate' access to other cues may be necessary (see Van Dyke and McElree 2011). In such a model, some cues would be given priority over others in determining probability of retrieval, or matching on a particular cue would be a precondition to being considered for retrieval at all. Under such a framework, c-command information might serve as a 'gating' cue. If this were the case a non-c-commanding noun phrase's morphological features should not interfere with antecedent retrieval.

*Previous Work On Bound Pronouns*

Several prior studies on bound pronouns have investigated whether bound variable readings are preferred over coreferential readings for pronouns in context (Avrutin, 1994; Shapiro and Hestvik, 1995; Foley, et al. 1997; Guo, et al. 1996; Frazier and Clifton, 2000; Cunnings, et al. 2011). Because the studies are concerned with preferences that comprehenders exhibit in the post-retrieval resolution process of pronominal interpretation they are largely orthogonal to our purposes. The two questions of central interest in this paper have received less attention in the previous psycholinguistics literature.

Many syntactic studies have presented contrasts that motivate the c-command constraint on bound-variable pronoun licensing (e.g. Reinhart, 1983; Büring, 2005). The empirical purchase of the generalization can be observed across a wide range of constructions. For example, a quantified noun phrase cannot bind a pronoun from within a relative clause (5), or across a
sentential conjunct (6). Co-reference between a pronoun and a referential noun phrase in the same position is nevertheless acceptable, because coreference is not subject to the same c-command constraint.

(5) The spectators that couldn't find the/*any cyclist continued to look for him.

(6) Mary said Jane met the/*every cyclist, but she said that Felicia didn’t like him.

Some studies have tested the acceptability of bound variable pronoun relations in specific constructions to argue that the c-command generalization should be weakened or that it is not uniform. As part of a larger investigation of the effects of c-command and linear order on the acceptability of pronoun-antecedent relations Gordon and Hendrick (1998) conducted a series of offline judgment study to determine the role of c-command in establishing QP-pronoun binding dependencies. The authors found that c-command conferred a benefit in acceptability to the bound-variable pronoun relation, but found that participants nevertheless rated sentences like (7), where a feature-matching antecedent preceded the pronoun, as significantly more acceptable than sentences in which the antecedent-pronoun order was reversed. This led the authors to argue in favor of de-emphasizing the role of strict c-command in determining antecedent acceptability.

(7) Each girl’s parents decided what she could do.

Barker (2012) relatedly cited the relative acceptability of cases like (7) as evidence against a c-command constraint on variable-binding. Other cases of apparently acceptable bound-variable pronoun interpretations occur in cases of inverse linking (8), or in cases where a quantified
phrase may bind a pronoun from inside a prepositional phrase (9).

(8) Someone from every city hates it. (May, 1985)

(9) A book was given to every boy by his mother. (Harley, 2003:64)

The relative acceptability of these constructions suggests that a class of acceptable bound-variable pronoun dependencies might better be understood in terms of non-surface command (i.e., quantifier scope or logical form LF c-command – see May, 1977, or Kush, 2013 for discussion). Either way, whether it is surface or non-surface command relations that best characterize the distribution of licit binding relations, structural restrictions (i.e. the relative position of the QP with respect to the pronoun) play a deciding role in determining whether a QP may antecede a pronoun.

Other research has shown that some instances of QP-pronoun dependencies that do not obey the c-command constraint should be analyzed as instances of non-binding dependencies. These include e-type pronoun readings (10 - also referred to as donkey pronouns, see Heim 1982; Elbourne, 2005) and instances of Telescoping (11), in which a pronoun appears to co-vary in interpretation with a quantified phrase across a sentence boundary (Roberts, 1989, Poesio & Zucchi, 1992; Anderssen, 2011, inter alia).

(10) Every farmer who owns a donkey beats it.

(11) Every degree student walked up to the stage. He collected his diploma and then sat down.

Among the ways that these constructions differ from true cases of bound-variable pronoun
dependencies in that the quantifiers that participate in the constructions are more sharply limited than in binding, they pattern with instances of coreference with respect to substitution (Büring, 2005), and they are subject to semantic and pragmatic constraints not observed on other binding dependencies (see Poesio and Zucchi, 1992).

As part of a larger series of studies intending to investigate the role of c-command in licensing bound-variable pronouns, Carminati, Frazier and Rayner (2002) conducted an eye-tracking study that tested participants' ability to interpret a pronoun as dependent on a preceding feature-matching antecedent. The antecedent was either a c-commanding QP, a non-commanding QP, or a referential noun phrase. Carminati and colleagues’ Experiment 1 varied the REFERENTIALITY of a feature-matching phrase (Every/The British soldier), and the C-COMMAND relation between the phrase and a later pronoun he. The pronoun was either embedded in a subordinate clause c-commanded by the phrase (12), or placed in a separate conjunct, as in (12).

(12)  \textit{C-command}

\{Every/The\} British soldier thought that \textit{he} killed an enemy soldier.

(13)  \textit{No C-command}

\{Every/The\} British soldier aimed and then \textit{he} killed an enemy soldier.

The authors found that reread times at the pronoun in QP sentences were longer than in Referential sentences. The authors found no differences that suggested that participants exhibited more difficulty processing the pronoun in sentences with non-c-commanding QPs than in sentences with c-commanding QPs. The authors took these results to indicate that participants’ retrieved QPs equally easily for pronouns, regardless of c-command. Other researchers have also
adopted this interpretation. For example, Barker (2012) recently cited the study as evidence against the c-command constraint on bound-variable pronouns.

We do not interpret Carminati and colleagues' findings as counterevidence to the viability of a c-command constraint on bound-variable pronoun licensing for two reasons. First, the data do not clearly rule out that antecedents are not distinguished on the basis of c-command. Numerical trends suggested a pairwise difference between the quantificational conditions that were not observed between referential conditions. Reading times (first-pass and re-read times) were disproportionately higher in the condition in which c-command mediated binding was required to establish a sentence-internal relation between antecedent and pronoun. Second, the experimental materials exhibit characteristics of Telescoping constructions (e.g., they do not support downward-entailing quantifiers such as no and are subject to a number of other semantic restrictions). As instances of Telescoping, the materials are therefore not relevant for determining the c-command generalization on true instances of bound variable pronouns (see Kush, 2013, Chapter 7 for more discussion). The same critique extends to the remaining experiments in Carminati et al. (2002).

In sum, there is only limited research to date that formally tests acceptability of c-command generalization on bound-variable pronoun interpretations and almost none that tests whether the c-command generalization has online effects.

The second question of interest is whether online antecedent retrieval is able to selectively ignore antecedents that are judged to be unacceptable in offline measures. Little is known on this topic.
Once again, although Carminati and colleagues' findings are cited as evidence that non-c-commanding QPs are accessible antecedents under some circumstances, the studies only show that non-c-commanding QPs are accessible when Telescoping is grammatically possible and therefore acceptable.

In what follows we present two experiments that address each of our questions of interest in turn. Experiment 1 tests whether antecedent retrieval can distinguish between grammatically appropriate from inappropriate quantified phrases based on their c-command relation to a pronoun. Experiment 2 tests whether antecedent retrieval can selectively ignore non-c-commanding QPs.

**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 the c-command relation facilitates linking a pronoun to a quantificational antecedent in a manner that it does not affect establishing a link to a referential antecedent. With regard to the key questions outlined above, the current study was designed to test whether antecedent access is sensitive to c-command relations, but could not test whether antecedent access mechanisms are sufficiently sensitive to ignore non c-commanding QPs. The experiment used materials that were designed to facilitate bound variable interpretations for the pronoun, while excluding the possibility of Telescoping or other related interpretations.

**Materials**
The experiment used a $2 \times 2$ 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 a conjunction or a connective. 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 that mismatched the gender of the pronoun and was followed by a second noun phrase that matched the gender of the pronoun, the *antecedent*. Potential antecedents were placed in an embedded clause to prevent any advantage in retrieval probability associated with first-mention, or being the main subject of the sentence (Corbett and Chang, 1983). A subsequent noun phrase in the first clause included a possessive pronoun 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.

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 attachment height of the clause containing the pronoun, we manipulated whether a c-command relation holds between the pronoun and feature-matching NP, without changing the position of the feature-matching NP.

The conditions compared the effect of c-command as manipulated by **STRUCTURE** on the accessibility of feature-matching *Referential* and *Quantificational* noun phrases in the same structural positions in order to test retrieval's sensitivity to the c-command constraint on variable
binding because the accessibility of referential antecedents is not grammatically determined by 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. We chose to use any, rather than quantifiers such as every or each that have been used in previous studies (e.g., Carminati et al., 2002), because downward-entailing quantifiers such as any or no have a more constrained distribution as pronoun antecedents and in quantifier scope and do not license the additional interpretations that are sometimes found with upward-entailing quantifiers such as every (see, e.g., Beghelli & Stowell, 1997). 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.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantificational-But</td>
<td>Kathi didn’t think any janitor liked performing his/ custodial duties,/</td>
</tr>
<tr>
<td></td>
<td>but he/ had to/ clean up messes/ left after prom anyway.</td>
</tr>
<tr>
<td>Quantificational-When</td>
<td>Kathi didn’t think any janitor liked performing his/ custodial duties/</td>
</tr>
<tr>
<td></td>
<td>when he/ had to/ clean up messes/ left after prom.</td>
</tr>
<tr>
<td>Referential-But</td>
<td>Kathi didn’t think the janitor liked performing his/ custodial duties,/</td>
</tr>
<tr>
<td></td>
<td>but he/ had to/ clean up messes/ left after prom anyway.</td>
</tr>
<tr>
<td>Referential-When</td>
<td>Kathi didn’t think the janitor liked performing his/ custodial duties / when he/ had to/ clean up messes/ left after prom.</td>
</tr>
</tbody>
</table>

Table 1. Example Items for Experiment 1. Slashes indicate regions of analysis used for Experiment 1c.
We predicted that feature-matching referential antecedents should always be accessible to retrieval, irrespective of the connective, because of the relatively liberal constraints on antecedents for referential pronouns. In contrast, we predicted that the choice of connective should impact the linking of a pronoun to a quantificational antecedent. In the *Quantificational-When* condition, the feature-matching QP c-commands the pronoun, and hence should be readily accessible as an antecedent for the pronoun. But in the *Quantificational-But* condition the QP fails to c-command the pronoun, and hence it should be less available as an antecedent for the pronoun, leading to increased processing difficulty. The increased difficulty could reflect difficulty in linking the pronoun to the QP, or it could reflect the parser’s attempt to find an alternative antecedent. Therefore, if retrieval of potential antecedents is sensitive to relational
information, a REFERENTIALITY × STRUCTURE interaction is predicted. Alternatively, if retrieval of antecedents is not sensitive to relational constraints on bound variable antecedents, then we expect no interaction, although main effects of antecedent type or structure are possible, reflecting independent effects of resolving referential and bound variable interpretations or of the accessibility of different structural positions.

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 the structural generalization in offline ratings, using two measures: simple acceptability ratings and a paraphrase selection task as a measure of interpretation. These offline tests are important for establishing whether the non-c-commanding QPs in the experimental materials do not participate in telescoping or otherwise grammatically-mediated dependencies with the pronoun.

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, developed and managed by Alex Drummond (www.spellout.net/ibexfarm/docs). 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 endpoints 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.

Materials
24 sets of four experimental items as in (18) were distributed across 4 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. 18 of the unacceptable fillers contained errors such as unlicensed negative polarity items, subject-verb agreement errors, or were missing function words such as the. 8 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 base-line against which to compare the acceptability of the Quantificational-But conditions. 6 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. 4 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, 4 sentences featuring an infelicitous pronoun and the connectives but or and were included for the same reason. 24 additional acceptable fillers were also included.

<table>
<thead>
<tr>
<th></th>
<th>Good-Binding</th>
<th>Bad-Binding</th>
<th>Bad-Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fillers</td>
<td>5.82 (.16)</td>
<td>4.28 (.17)</td>
<td>3.05 (.11)</td>
</tr>
</tbody>
</table>

Table 2. Mean acceptability judgments from Experiment 1a. Standard errors in parentheses.

**Results**

Mean acceptability judgments are provided in Table 2. A reliable effect of REFERENTIALITY was observed ($\beta = -0.60$, s.e. = .08, $t = -7.71$, $p < .001$), driven by higher scores in the Referential conditions than in the Quantificational conditions. A marginally significant effect of STRUCTURE was observed ($\beta = -0.13$, s.e. = 0.08, $t = -1.76$, $p < .10$), due to higher acceptability scores in the When conditions. The REFERENTIALITY $\times$ STRUCTURE interaction was significant ($\beta = -0.18$, s.e. = .08, $t = -2.35$), with Quantificational-But conditions receiving the lowest acceptability scores overall. The pairwise difference between the two Quantificational conditions was significant ($t = -2.86$, $p < .01$). Quantificational-When sentences were rated reliably more acceptable than Quantificational-But sentences. No comparable pairwise difference was observed between
Referential conditions ($t < 1$).

Because control sentences were not factorially manipulated we do 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.

**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.

Sentences in which the feature-matching antecedent was referential were given ratings similar to each other irrespective of the c-command status between antecedent and pronoun. Sentences with a quantificational antecedent were rated, on average, as less acceptable than their referential counterparts. The acceptability of sentences with a quantificational antecedent differed as a function of the QP's c-command relation to the pronoun. Sentences in which the QP antecedent c-commanded the pronoun were rated as significantly more acceptable than those in which the antecedent did not c-command the pronoun.

The qualitative pattern of acceptability scores observed is consistent with the generalization from the linguistics literature that c-command is a determinant of a QP's accessibility as an antecedent for a pronoun. However, the relatively high acceptability scores in the *Quantificational-But* condition (relative to the ungrammatical fillers) are unexpected on the assumption that a pronoun
that lacks an overt antecedent should have induced unacceptability. This relatively high acceptability is consistent with two interpretations. First, it might suggest that c-command influences the acceptability of a QP-pronoun binding dependency, but that it does not exhaustively determine the possibility 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 slight cost. The second interpretation of these results is that the pronoun was interpreted as disjoint from the QP in accordance with the grammatical generalization, but that accommodation of an antecedent-less pronoun in offline judgment studies may not always result in dramatic reductions in acceptability as was initially assumed. This second interpretation receives some suggestive support from the similarly elevated acceptability ratings of the Bad-Binding condition, which also contained a pronoun without an overt antecedent. 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 the relatively high acceptability ratings for the Quantificational-But condition in the study above reflects that participants do entertain a binding dependency between a QP and pronoun in the absence of c-command, contrary to the generalizations in the formal linguistics literature, we conducted a forced-choice interpretive task to measure participants’ consideration of bound-variable interpretations.

Participants
21 participants were recruited through Amazon Mechanical Turk. Participants were paid $3.50 for their participation.

*Materials and Procedure*

The interpretation task used the same target sentences as the acceptability rating task. A target sentence was drawn from the items from the $2 \times 2$ design above. Each test sentence was paired with two corresponding response sentences to form a triplet. Participants’ choice of the paraphrase that best matched their interpretation of the test sentence indicated their interpretation of the pronoun in the second clause of the target sentence. These response sentences were designed to be either consistent with a *single-individual* interpretation (response sentences marked *SI*), or a *quantificational* interpretation of the continuation (marked *BV*). 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 response sentences differed minimally based on the semantics of the test sentence. Examples are given below.

(13) **Example But Sentence**

Kathi didn’t think the 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.

**BV.** Every janitor had to clean up after prom.
(14) **Example When Sentence**

Kathi didn’t think any janitor enjoyed his custodial duties when he had to clean up messes left after prom anyway.

**SI.** There was someone who disliked having to clean up after prom.

**BV.** 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 saw a screen on which the sentence triplet was center-aligned. Response sentences were presented as a numbered list in randomized order. After choosing a sentence, participants were asked to rate their confidence in their decision.

Test sentences were interspersed randomly among 28 filler triplets, which themselves were consistent either with single-individual, or quantificational interpretations. Participants were encouraged to answer as accurately as possible and received feedback on their accuracy throughout the experiment 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 four participants who scored lower than 70% accuracy on either the filler questions or the single-individual test sentences were excluded from analysis. Statistical analysis used logistic-mixed effect regression. **STRUCTURE** and **ANTECEDENT TYPE** were treated as fixed
effects. Subject and item were treated as random effects.

Results

The proportion of trials, by condition, on which participants chose the sentence corresponding to the quantificational interpretation of the continuation is in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>When</th>
<th>But</th>
<th>Filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantificational</td>
<td>0.90</td>
<td>0.32</td>
<td>0.89</td>
</tr>
<tr>
<td>Referential</td>
<td>0.06</td>
<td>0.08</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 3. Proportion of Quantificational Paraphrases Chosen by Participants in Experiment 1b.

Statistical analysis revealed that a main effect of STRUCTURE was reliable (β = -3.37, s.e. = 0.43, z = -7.84, p < .001). This was driven by an increased proportion of quantificational responses in the Quantificational-When condition in comparison to the Quantificational-But condition. A main effect of REFERENTIALITY was also significant (β = -2.21, s.e. = 0.51, z = -4.37, p < .001), driven by an increased proportion of quantificational responses in Quantificational conditions relative to Referential conditions. A significant STRUCTURE × ANTECEDENT TYPE interaction was also observed (β = -3.05, s.e. = 0.74, z = -4.13, p < .001), due to a higher proportion of quantificational responses in the Quantificational-When condition relative to all other conditions.

A main effect of STRUCTURE was also observed in the pairwise comparison between Quantificational and Single filler conditions (β = -3.40, s.e. = 0.26, z = -12.96, p < .001). Filler-Quantificational sentences were more consistently given a quantificational interpretation than Filler-Single sentences.
Discussion

Experiment 1b probed participants' preferred interpretations for pronouns in sentences containing a feature-matching potential antecedent by having them choose paraphrases of test sentences. Paraphrases were provided in which the pronoun was interpreted as referential or as quantificationally bound. The c-command relation between the antecedent and the pronoun was manipulated, as was the quantificational status of the potential antecedent. When the potential antecedent was referential, the pronoun and potential antecedent are able to enter into a coreference relation, irrespective of their c-command relation. Participants were predicted to choose referential paraphrases under these conditions. When the potential antecedent was quantificational, a grammatical binding relationship was only possible in the condition where the QP c-commanded the pronoun. Participants' consideration of the bound interpretation in both conditions was measured as the proportion of trials where they chose the quantificational paraphrase.

In filler conditions, participants reliably picked the correct interpretation. In 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.

The proportion of quantificational readings in the Quantificational-When condition was comparable to the proportion of quantificational readings chosen in Quantificational fillers. This indicates that participants chose to form quantificational binding dependencies for feature-
matching pronouns in the sentence, rather than positing an extra-sentential referent for the pronoun.

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 70% of trials participants favored the relatively costly strategy of coercing an unknown referent for resolution of the pronoun, rather than considering a grammatically illicit bound reading using the intra-sentential QP as the pronoun’s antecedent. Individual participant analyses show that most participants chose the quantificational interpretation for Quantificational-But trials at very low rates. 11 of the 17 participants chose the quantificational interpretation on 30% or fewer trials (average proportion 0.16). In contrast, the remaining 6 participants chose the quantificational response at much higher rates (range 0.50 – 1.00). Informal inspection of the data suggests that at least some of these participants exhibited an overall bias towards quantificational responses.

The findings demonstrate a clear effect of Structure, but only in the Quantificational conditions, where the structural manipulation affects grammatical availability of a bound variable reading. Whether these results reflect a grammatical distinction or simply an interpretive preference cannot conclusively be determined because the task only probed participants’ interpretations. Regardless of the origin of the decision process, the results show that participants do make the distinction between pronouns that can and cannot be bound by a QP. These results also suggest that the relatively high ratings of Quantificational-But sentences in Experiment 1a do not reflect participants choosing a bound variable reading for the pronoun. The results are instead more consistent with an interpretation under which coercion of an extra-sentential referential
antecedent for the pronoun in *Quantificational-But* conditions does result in drastic reductions in acceptability. Having validated the ability of the structural manipulation to affect the acceptability of quantificational phrases as antecedents, Experiment 1c tests whether real-time indices of retrieval reflect this distinction.

*Experiment 1c. Eye-tracking while reading*

Having established participants' offline sensitivity to the grammatical generalization that a QP's c-command relation in relation to a pronoun affects its eligibility as an antecedent, 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*

44 participants were recruited from the University of Maryland community (26 female, mean age 21.2). Participants received course credit or monetary compensation. Paid 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.

*Materials & 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. Each character was 9x16 pixels on the display. The resolution of the LCD visual display was 1280 x 720 pixels. 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 1000Hz. 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 manipulated connective and the critical pronoun. The connective was included in the critical pronoun region due to the tendency towards pronouns being skipped at high rates due to their relatively short length (Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994). The post-pronoun region contained the first three words of the verb phrase that followed the pronoun. The remainder of each sentence was treated as a single region.

We present data from four regions: pre-critical, critical pronoun, post-pronoun and final regions.

For each region of interest we report four different measures. These measures can be divided into two categories: early measures and late measures. Early measures reported here are first-pass reading time and right-bound reading time. 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. The late measures reported here are reread time and total time. Reread times are calculated by summing all fixations that occur in a region after a participant has first exited the region to the right. Total times sum over all fixations in a particular region of interest, including first pass reading time and any time spent rereading the region. For all late measures, trials on which a region was not re-fixated contributed a value of 0ms to the cell mean.
Statistical analysis was carried out using 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*), **Continuation** (whether the critical pronoun was preceded by *but*, or *when*), and their interaction, as well as random intercepts for participants and items (Baayen, Davidson & Bates 2008), together with participant and item random slopes for all fixed effects (Barr, Levy, Scheepers & Tily, 2013). In adopting a mixed-effect models analysis, missing values were left out from analysis, rather than being replaced as zeros. Analyses were also conducted on data in which missing observations contributed a value of 0ms to the cell mean. Analyses were also computed with missing values simply omitted. Effects did not change between the two analysis regimes.

A fixed effect was considered significant if its absolute t-value was greater than 2, which indicates that its 95% confidence interval does not include 0 (Gelman & Hill, 2005). Reported coefficients whose absolute t-value is greater than 2 are significant at the p < 0.05 level.
**Results**

<table>
<thead>
<tr>
<th></th>
<th>Pre-Critical Region</th>
<th>Critical Pronoun</th>
<th>Post-Pronoun Region</th>
<th>Final Region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Pass</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quant-But</td>
<td>422 (16)</td>
<td>306 (12)</td>
<td>416 (17)</td>
<td>455 (18)</td>
</tr>
<tr>
<td>Quant-When</td>
<td>457 (18)</td>
<td>324 (11)</td>
<td>377 (14)</td>
<td>437 (16)</td>
</tr>
<tr>
<td>Refer-But</td>
<td>420 (16)</td>
<td>313 (11)</td>
<td>400 (16)</td>
<td>443 (19)</td>
</tr>
<tr>
<td>Refer-When</td>
<td>442 (19)</td>
<td>316 (12)</td>
<td>411 (15)</td>
<td>410 (16)</td>
</tr>
<tr>
<td><strong>Right-bound</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quant-But</td>
<td>545 (19)</td>
<td>333 (14)</td>
<td>474 (21)</td>
<td>505 (20)</td>
</tr>
<tr>
<td>Quant-When</td>
<td>570 (21)</td>
<td>357 (13)</td>
<td>431 (17)</td>
<td>492 (17)</td>
</tr>
<tr>
<td>Refer-But</td>
<td>557 (21)</td>
<td>350 (16)</td>
<td>454 (18)</td>
<td>493 (21)</td>
</tr>
<tr>
<td>Refer-When</td>
<td>538 (21)</td>
<td>343 (13)</td>
<td>472 (16)</td>
<td>456 (17)</td>
</tr>
<tr>
<td><strong>Reread</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quant-But</td>
<td>519 (33)</td>
<td>384 (28)</td>
<td>469 (34)</td>
<td>486 (35)</td>
</tr>
<tr>
<td>Quant-When</td>
<td>698 (49)</td>
<td>445 (29)</td>
<td>518 (36)</td>
<td>542 (35)</td>
</tr>
<tr>
<td>Refer-But</td>
<td>567 (41)</td>
<td>390 (28)</td>
<td>501 (34)</td>
<td>457 (28)</td>
</tr>
<tr>
<td>Refer-When</td>
<td>536 (34)</td>
<td>426 (27)</td>
<td>466 (25)</td>
<td>523 (37)</td>
</tr>
<tr>
<td><strong>Total Times</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quant-But</td>
<td>701 (28)</td>
<td>458 (19)</td>
<td>662 (28)</td>
<td>680 (29)</td>
</tr>
<tr>
<td>Quant-When</td>
<td>831 (38)</td>
<td>532 (22)</td>
<td>648 (29)</td>
<td>731 (29)</td>
</tr>
<tr>
<td>Refer-But</td>
<td>710 (28)</td>
<td>478 (21)</td>
<td>652 (27)</td>
<td>657 (27)</td>
</tr>
<tr>
<td>Refer-When</td>
<td>722 (28)</td>
<td>507 (21)</td>
<td>633 (24)</td>
<td>670 (27)</td>
</tr>
</tbody>
</table>

Table 4. Mean raw reading times by measure and region for Experiment 1c. Standard errors in parentheses.
Figure 3. (a) Rightbound reading times for the post-pronoun region and, (b) Total reading times in the pre-pronoun region in Experiment 1c. Error bars indicate one standard error of the mean.

Pre-critical Region. In the pre-critical region, a main effect of STRUCTURE was marginally significant in first-pass times ($\beta = -27.0$, s.e. = 14.4, $t = -1.88$, $p < .10$). This reflected longer reading times in *When* conditions relative to *But* conditions. No effects achieved significance in right-bound times. The ANTECEDENT TYPE $\times$ STRUCTURE interaction was significant ($\beta = 165.0$, s.e. = 90.3, $t = -1.82$, $p < .10$), due to longer RTs in the *Quantificational-When* condition than in the other three conditions.
In total time, a main effect of Antecedent Type was marginally significant \((\beta = 47.2, \text{s.e.} = 30.0, t = 1.57)\) and a main effect of Structure was significant \((\beta = -68.9, \text{s.e.} = 30.4, t = -2.26, \ p < .05)\). The interaction was significant, as well \((\beta = -121.0 \text{ s.e.} = 55.4, t = -2.19, p < .05)\). In total times the interaction was driven by elevated RTs in the Quantificational-When condition only. Pairwise comparison between the Quantificational conditions was significant \((\beta = -128.80 \text{ s.e.} = 47.01, t = -2.74, p < .05)\), but not between Referential conditions \((t < 1)\). The interaction in total reading times in the pre-pronoun region suggests that the parser distinguishes between the two Quantificational conditions.

**Critical Pronoun Region.** No significant main or interaction effects were observed in any measures at the critical pronoun region.

**Post-Pronoun Region.** At the post-pronoun region, there was a marginally significant Antecedent Type \(\times\) Structure interaction \((\beta = 49.2, \text{s.e.} = 28.4, t = 1.73, p < .10)\) in first-pass times. The interaction was driven by faster RTs in the Quantificational-When condition relative to the Quantificational-But condition. This pairwise difference was significant \((\beta = 43.5, \text{s.e.} = 18.5, t = 2.36, p < .05)\). No corresponding pairwise difference was observed between Referential conditions \((t < 1)\).

The Antecedent Type \(\times\) Structure interaction was also marginally significant in right-bound times \((\beta = 59.4, \text{s.e.} = 30.6, t = 1.94, p < .10)\). Pairwise comparison confirmed that RTs in the Quantificational-But condition were significantly longer in comparison to the Quantificational-When condition \((\beta = 45.8, \text{s.e.} = 21.3, t = 2.15, p < .05)\), but no differences were observed.
between the *Referential* conditions (*t* < 1).

**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 was able to vary the structural relation between the antecedent and the pronoun while holding linear distance constant by means of 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 only QPs that c-command a pronoun may bind it. 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 participants reliably choose to interpret pronouns as bound by a QP only when the QP c-commands the pronoun.
Eye-tracking measures showed that antecedent retrieval distinguished structurally-appropriate QPs from inappropriate QPs. This distinction was manifested in two patterns in the reading time measures. In early measures (first-pass and right-bound time), the Quantificational-When condition was read significantly more quickly than its Quantificational-But counterpart in the region immediately following the critical pronoun. No pairwise difference was observed between the Referential conditions in this region. In later measures (reread and total times), a different pattern of effects emerged. Longer RTs were observed in the Quantificational-When condition in the pre-critical pronoun region relative to all other conditions.

These results support the contention that pronoun resolution distinguishes potential QP antecedents based on the structural relation they bear to the critical pronoun. The interaction effect indicating sensitivity to c-command was observed in early measures, immediately following the critical pronoun. This slight delay in effects is typical in studies where the critical region is a short, high frequency word that is often skipped (e.g. Ehrlich & Rayner, 1983; Garrod, Freudenthal, & Boyle, 1994).

Participants displayed no difficulty in accessing a c-commanding quantificational antecedent (as evidenced by relatively short early measures in the Quantificational-When condition). Longer reading times in the Quantificational-But condition are consistent with the parser encountering difficulty at the pronoun because it must coerce a referent for the pronoun.

Longer reading times were observed in later measures in the Quantificational-When condition relative to the other conditions. We suggest that these effects should not be interpreted as indices
of retrieval difficulty. Instead, the increased reading time in late measures may reflect the additional computational cost of processing a bound quantificational interpretation during the resolution stage following successful (unimpeded) retrieval. Previous studies have shown that establishing a licit dependency between a quantificational antecedent and a pronoun results in longer reading times in comparison to the time required to establish a coreferential dependency (e.g. Carminati et al. 2002).

If the resolution of a pronoun with a referential antecedent results in shorter reading times in late measures than the resolution of a quantificationally bound pronoun, the pronoun in the Quantificational-But displays the RT hallmarks of a referential pronoun. With respect to reread and total times the Quantificational-But condition is comparable to the two Referential conditions. We conclude that a bound variable reading is not considered for the pronoun in that condition. Rather, it is treated as referential. This aligns with the judgment patterns observed in Experiment 1b, where participants consistently chose to interpret the pronoun as a referential pronoun rather than a bound pronoun.

The results of Experiment 1 establish that antecedent retrieval can distinguish grammatically appropriate QPs from grammatically inappropriate QPs. The study does not determine whether this information categorically ‘gates’ access to QPs or whether such information acts as one retrieval cue among many other cues (e.g. morphological feature-match) in determining the accessibility of an item to retrieval. Experiment 1 cannot distinguish between these two hypotheses because potential antecedents were identical in their feature composition except for their c-command information. 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 can distinguish between c-commanding QPs that are eligible antecedents for a pronoun and non-c-commanding QPs that are grammatically inappropriate to antecede 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 $2 \times 2$ factorial design and the remaining 2 conditions were lexically similar, though structurally distinct, control conditions. An example is provided in Table 5.

**Test Conditions**

The test conditions used a $2 \times 2$ factorial design, which crossed the factors GENDER MATCH and REFERENTIALITY. All test items were transitive sentences with a plural definite subject. A singular pronoun that required an antecedent appeared as the object of the sentence's 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 definiteness were manipulated. The position of the potential antecedent was held
constant in Experiment 2 so the potential antecedent never c-commanded the pronoun.

The factor GENDERMATCH manipulated the whether the potential antecedent matched or mismatched pronoun in gender features by exchanging the potential antecedent's head noun. Strongly gender-biased nouns (e.g. doctor), 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 & Trofe (2003). Pronoun gender was held constant across test items and the number of masculine and feminine pronouns was counter-balanced across items.

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 downward-entailing 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 pronoun. Thus, we predicted undisrupted processing of the pronoun in 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. Gerrig, 1986; Osterhout & Mobley, 1995; van Hoek, 1997; Filik, Sanford & Leuthold, 2008). In both Referential-Mismatch and Quantificational-Mismatch conditions, indices of difficulty were expected at the pronoun. In the Quantificational-Match condition, the pronoun matched the
potential antecedent in gender features, but the antecedent was not a grammatical antecedent because it did not c-command the pronoun. Based on the findings from the previous experiment, we expected this lack of c-command to manifest as increased difficulty in processing the pronoun relative to the Referential-Match condition. The previous experiment did not establish whether the ability to distinguish between c-commanding and non-c-commanding QPs reflected that inappropriate QPs were completely inaccessible to retrieval, or that such QPs were simply less activated but nevertheless accessible. If the c-command status of the quantificational potential antecedent makes it inaccessible to retrieval, the difficulty observed in the Quantificational-Match condition should be comparable to the difficulty observed in the Mismatch conditions, in which no antecedent for the pronoun is found. On the other hand, if the c-command status of merely lowers availability of the QP but does not restrict access altogether, we expected to observe evidence of facilitatory interference. Processing difficulty should be lower in Quantificational-Match conditions relative to Mismatch conditions, reflecting mis-retrieval of the inappropriate QP based on its gender features. This type of interference (also called intrusive licensing) has been observed in other retrieval-mediated dependencies such as negative polarity item licensing (e.g. Vasishth et al., 2008) and in cases of agreement attraction.

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; Lakoff, 1970; Rodman, 1976; Sharvit, 1999, a.o.). In the sentence The woman that every man loves is his mother the QP every man can apparently bind the pronoun his, producing a functional 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 (cf. Heycock, 1991, 1992). Our materials do not license functional readings because they are not copular sentences. Moreover, our use of direct object pronouns makes functional readings less likely because the readings appear to favor pronouns that are possessors. Finally, although Sharvit (1999) has argued that functional readings can arise with non-copular sentences, these deviations from the generalization are only observed with the quantifiers every, and each. Our use of no rules out the possibility that the parser considers a functional reading (consider the unacceptability of *The woman that no man wanted to marry talked with his mother).
(e.g., Wagers, Lau & Phillips, 2009; Dillon et al., 2013).

**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 tri-clausal sentences in which the first verb was a propositional attitude verb whose subject was a plural NP. 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 QP and pronoun.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referential-Match</td>
<td>The troop leaders that the girl scout had no respect/ for had scolded her/ after the/ incident at scout camp./</td>
</tr>
<tr>
<td>Referential-Mismatch</td>
<td>The troop leaders that the boy scout had no respect/ for had scolded her/ after the/ incident at scout camp./</td>
</tr>
<tr>
<td>Quantificational-Match</td>
<td>The troop leaders that no girl scout had respect/ for had scolded her/ after the/ incident at scout camp./</td>
</tr>
<tr>
<td>Quantificational-Mismatch</td>
<td>The troop leaders that no boy scout had respect/ for had scolded her/ after the/ incident at scout camp./</td>
</tr>
<tr>
<td>Control-Match</td>
<td>The troop leaders were sure no girl scout/ was afraid that she/ would be scolded/ after the incident at scout camp.</td>
</tr>
<tr>
<td>Control-Mismatch</td>
<td>The troop leaders were sure no boy scout/ was afraid that she/ would be scolded/ after the incident at scout camp.</td>
</tr>
</tbody>
</table>

Table 5. Example Test and Control Items for Experiment 2. Regions of Analysis for Experiment 2b indicated by slashes.
Experiment 2a: Acceptability Judgment Study

To first verify offline sensitivity to the c-command constraint with Experiment 2’s materials, an online acceptability judgment study was run.

Participants

20 participants were recruited through Amazon Mechanical Turk marketplace (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. 30 of the 36 items used in the eye-tracking study were used in the judgment study. 4 sentences involving the same configuration as the control conditions, but with referential embedded subjects, were also included (Referential-Command sentences). 6 additional sentences involving variable binding in various configurations were included in the list to provide an index of the general availability/acceptability of variable binding (Good-Binding sentences). 58 filler sentences, all grammatical and of comparable complexity, were also included. This resulted in the presented sentences being roughly 70% acceptable.
Results

Average acceptability judgment ratings for Experiment 2a are given in Table 6. Within the Test conditions, a main effect of Referentiality was significant ($\beta = -1.00$, s.e. = 0.148, $t = -6.75$, $p < .001$), due to higher average acceptability ratings for Referential sentences over Quantificational sentences. A reliable effect of GenderMatch was also observed ($\beta = -1.16$, s.e. = 0.148, $t = -7.84$, $p < .001$). Match conditions were rated more acceptable than Mismatch conditions. The Referentiality $\times$ GenderMatch interaction was also significant ($\beta = 1.59$, s.e. = 0.297, $t = 5.36$, $p < .001$). The Referential-Match condition received the highest mean acceptability rating, while all other conditions were rated relatively low. A pairwise difference between Quantificational conditions was also observed. The Quantificational-Match condition was rated slightly, but significantly, higher than the Quantificational-Mismatch condition ($\beta = 0.37$, s.e. = 0.176, $t = -2.11$, $p < .05$).

Pairwise comparison of the two control conditions revealed a significant effect of GenderMatch ($\beta = 1.78$, s.e. = 0.244, $t = 7.29$, $p < .001$). The Control-Match condition received significantly

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average Acceptability Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referential-Match</td>
<td>4.14 (.20)</td>
</tr>
<tr>
<td>Referential-Mismatch</td>
<td>2.18 (.16)</td>
</tr>
<tr>
<td>Quantificational-Match</td>
<td>2.35 (.17)</td>
</tr>
<tr>
<td>Quantificational-Mismatch</td>
<td>1.97 (.15)</td>
</tr>
<tr>
<td>Control-Match</td>
<td>3.97 (.20)</td>
</tr>
<tr>
<td>Control-Mismatch</td>
<td>2.19 (.14)</td>
</tr>
<tr>
<td>Filler: Referential-Command</td>
<td>4.66 (.23)</td>
</tr>
<tr>
<td>Filler: Good-Binding</td>
<td>5.40 (.15)</td>
</tr>
</tbody>
</table>

Table 6. Average acceptability ratings for items in Experiment 2. Standard errors in parentheses.
higher acceptability ratings than the *Control-Mismatch* condition.

**Discussion**

The study sought to validate the grammatical generalization that non-c-commanding referential NPs are acceptable antecedents for a pronoun, but QPs that do not c-command a pronoun are not eligible antecedents.

The filler condition whose structures matched the syntactic configuration used in *Control* sentences received a high average acceptability score, which suggests that participants had no trouble with the configurations *per se*. Participants also gave high average acceptability scores to fillers that were uncontroversial cases of grammatical binding. This confirms that participants found bound variable readings available and acceptable.

Test conditions that lacked a feature-matching antecedent received low acceptability ratings. The condition in which a feature-matching referential antecedent was available to the pronoun exhibited the highest average acceptability. Similar ratings of acceptability were observed in the *Control-Match* condition, where a matching QP c-commanded a subject pronoun in the lower clause. The *Quantificational-Match* condition, in which the potential antecedent matched the pronoun in gender features but failed to c-command the pronoun, received lower ratings. These findings are consistent with the results of Experiment 1 that suggested grammatically appropriate feature-matching phrases were more accessible as antecedents than grammatically inappropriate phrases.

One effect emerged that was of direct relevance to the question of partial-match interference
from structurally inappropriate QPs. The condition in which the structurally inappropriate QP antecedent matches the pronoun in gender was rated slightly more acceptable than the condition in which the QP did not match the pronoun. This effect can be interpreted in one of two ways. First, the effect could indicate a lingering effect of retrieval interference. Mis-retrieval of the inappropriate QP as an antecedent for the pronoun may have occasionally caused participants to judge the pronoun as felicitous, rather than lacking an antecedent. The second option is that the effect reflects not interference, but rather some minor facilitation of a later interpretive process. Perhaps the presence of a structurally inappropriate matching QP facilitates the coercion of an extra-sentential referent for the pronoun.

Experiment 2b: Eye-tracking while Reading

Participants

30 participants from the University of Maryland community participated in the study for pay or course credit (18 females, mean age 20.8). Paid 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 for a total of 150 sentences per participant. 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 5. For the test conditions, the critical region contains the critical pronoun and the three characters that precede it. The three characters preceding the pronoun were included in the region to (i) account for possible parafoveal processing of the pronoun when reading the preceding context, and (ii) to increase the number of observations for the critical region. Small, high frequency words are often skipped in eye-tracking tasks (Rayner, 1997, 1998; Rayner et al., 2004; Rayner, Ashby, Pollatsek, & Reichle, 2004; Ashby, Rayner & Clifton, 2005; Staub & Rayner 2007 a.o.), therefore limiting the region to just the pronoun would result in a large amount of data loss. The post-pronoun region included the word or two words following the pronoun, depending on the length of the following prepositional phrase's head.

In Control sentences, the critical pronoun region was similarly delimited to the pronoun and the three characters that preceded it. In all sentences these three characters were the last three letters of the preceding complementizer. The post-pronoun region was defined as the entire passive verb phrase, excluding the PP adjunct which followed the verb and was considered as the sentence-final region.

Reading times reported are those reported in Experiment 1c. Statistical analysis was identical to that of Experiment 1c 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. The model also has a maximal random effects structure with random intercepts for subject and items, as well as random slopes for the fixed effects and their interaction by
participants and by items.

**Results**

<table>
<thead>
<tr>
<th>First-Pass</th>
<th>Pre-Critical Region</th>
<th>Critical Pronoun</th>
<th>Post-Pronoun Region</th>
<th>Final Region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quant-Match</strong></td>
<td>344 (13)</td>
<td>311 (13)</td>
<td>338 (27)</td>
<td>741 (38)</td>
</tr>
<tr>
<td><strong>Quant-Mismatch</strong></td>
<td>368 (16)</td>
<td>293 (13)</td>
<td>356 (24)</td>
<td>819 (46)</td>
</tr>
<tr>
<td><strong>Refer-Match</strong></td>
<td>364 (16)</td>
<td>254 (10)</td>
<td>294 (21)</td>
<td>848 (40)</td>
</tr>
<tr>
<td><strong>Refer-Mismatch</strong></td>
<td>358 (14)</td>
<td>285 (11)</td>
<td>318 (18)</td>
<td>803 (49)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Right-bound</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quant-Match</strong></td>
<td>399 (17)</td>
<td>342 (16)</td>
<td>403 (38)</td>
<td>1232 (61)</td>
</tr>
<tr>
<td><strong>Quant-Mismatch</strong></td>
<td>418 (19)</td>
<td>317 (14)</td>
<td>411 (32)</td>
<td>1383 (82)</td>
</tr>
<tr>
<td><strong>Refer-Match</strong></td>
<td>404 (17)</td>
<td>271 (12)</td>
<td>365 (39)</td>
<td>1280 (74)</td>
</tr>
<tr>
<td><strong>Refer-Mismatch</strong></td>
<td>396 (14)</td>
<td>330 (17)</td>
<td>417 (41)</td>
<td>1405 (70)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reread</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quant-Match</strong></td>
<td>518 (34)</td>
<td>410 (43)</td>
<td>398 (32)</td>
<td>1005 (86)</td>
</tr>
<tr>
<td><strong>Quant-Mismatch</strong></td>
<td>502 (48)</td>
<td>365 (28)</td>
<td>497 (59)</td>
<td>1268 (115)</td>
</tr>
<tr>
<td><strong>Refer-Match</strong></td>
<td>532 (49)</td>
<td>319 (34)</td>
<td>506 (79)</td>
<td>1133 (137)</td>
</tr>
<tr>
<td><strong>Refer-Mismatch</strong></td>
<td>558 (43)</td>
<td>455 (47)</td>
<td>502 (59)</td>
<td>1121 (95)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Time</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quant-Match</strong></td>
<td>585 (29)</td>
<td>438 (24)</td>
<td>527 (36)</td>
<td>1232 (61)</td>
</tr>
<tr>
<td><strong>Quant-Mismatch</strong></td>
<td>631 (33)</td>
<td>453 (23)</td>
<td>527 (35)</td>
<td>1428 (88)</td>
</tr>
<tr>
<td><strong>Refer-Match</strong></td>
<td>601 (33)</td>
<td>378 (22)</td>
<td>488 (42)</td>
<td>1280 (76)</td>
</tr>
<tr>
<td><strong>Refer-Mismatch</strong></td>
<td>673 (34)</td>
<td>473 (27)</td>
<td>533 (38)</td>
<td>1405 (70)</td>
</tr>
</tbody>
</table>

**Table 7.** Mean raw reading times by measure and region for test conditions in Experiment 2b. Standard errors in parentheses.
Figure 4. (a) First-pass reading times, (b) Right-bound reading times, and (c) Reread times at the pronoun region, Experiment 2b. Error bars indicate one standard error of the mean.
**Test Conditions**

**Pre-critical Region**

In the pre-critical region, no effects achieved significance in first-pass time, rightbound, or reread times (main effects ts < 1; interaction t < 1.6). In total time there was a marginally significant main effect of GENDERMATCH (β = 56.4, s.e. = 29.5, t = 1.91), driven by faster total times in Match conditions relative to their corresponding Mismatch conditions.

**Critical Pronoun Region**

In first-pass times, the model reveals a significant main effect of REFERENTIALITY (β = 31.6, s.e. = 14.0, t = 2.25, p < .05) driven by increased first-pass times in Quantificational conditions relative to Referential conditions. The REFERENTIALITY × GENDERMATCH interaction was marginally significant (β = -48.0, s.e. = 25.2, t= -1.91, p < .05), driven by lower RTs in the Referential-Match condition relative to the other three test conditions. Pairwise comparisons revealed a significant effect of REFERENTIALITY between Match conditions (β = 54.0, s.e. = 15.5, t = 3.49, p < .01), driven by longer RTs in the Quantificational-Match condition in comparison to the Referential-Match condition. No corresponding effect was observed between Mismatch conditions (t < 1). No pairwise effect of GENDERMATCH was found between Quantificational conditions (t < 1).

Effects similar to those observed in first-pass times were observed in right-bound times. A main effect of REFERENTIALITY was marginally significant (β = 31.7, s.e. = 17.9, t= 1.77), once again driven by longer reading times in Quantificational conditions. Pairwise comparisons revealed a significant effect of REFERENTIALITY within Match conditions (β = 69.2, s.e. = 18.4, t = 3.76, p < .01), but not within Mismatch conditions (t < 1). The REFERENTIALITY × GENDERMATCH interaction...
interaction was also significant ($\beta = -78.5$, s.e. = 32.8, $t = -2.40$, $p < .05$). Pairwise comparison between Referential conditions revealed a reliable effect of GENDERMATCH, due to increased reading times in the Referential-Mismatch condition ($\beta = -59.0$, s.e. = 20.8, $t = 2.84$, $p < .001$). Pairwise comparison between Quantificational conditions was not significant ($t < 1.2$), suggesting no reliable difference between the Quantificational-Match and Quantificational-Mismatch conditions.

In reread times, no significant main or interaction effects were found. In total times, there was a marginally significant main effect of GENDERMATCH ($\beta = -51.9$, s.e. = 26.1, $t = 1.99$). This effect was largely driven by a pairwise difference between Referential conditions ($\beta = 85.0$, s.e. = 28.7, $t = -2.97$, $p < .05$), with the Referential-Match condition read more quickly than the Referential-Mismatch condition. The pairwise difference between Quantificational conditions was not significant ($t < 1$).

Post-Pronoun Region
In the post-pronoun region there were no significant first-pass, right-bound, reread, or total Time effects ($ts < 1$).

Sentence-Final Region
In the sentence-final region, a marginal main effect of GENDERMATCH was found ($\beta = 206.0$, s.e. = 117.0, $t = 1.77$, $p < .10$) in reread times. This effect was driven by lower RTs in Match conditions, relative to Mismatch conditions. The marginal effect of GENDERMATCH was also observed in total time ($\beta = 118.0$, s.e. = 69.5, $t = 1.70$, $p < .10$), once again carried by lower RTs.
in GENDERMATCH conditions.

Control Conditions

Table 8. Raw mean reading times by measure and region for control conditions in Experiment 2b. Standard Error in Parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Pre-Critical Region</th>
<th>Critical Pronoun Region</th>
<th>Post-Pronoun Region</th>
<th>Pronoun+2 Region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Pass</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cont.-Match</td>
<td>775 (31)</td>
<td>309 (11)</td>
<td>250 (11)</td>
<td>471 (18)</td>
</tr>
<tr>
<td>Cont.-Mismatch</td>
<td>798 (36)</td>
<td>293 (11)</td>
<td>256 (9)</td>
<td>489 (20)</td>
</tr>
<tr>
<td><strong>Right-bound</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cont.-Match</td>
<td>924 (34)</td>
<td>326 (12)</td>
<td>262 (14)</td>
<td>533 (20)</td>
</tr>
<tr>
<td>Cont.-Mismatch</td>
<td>960 (43)</td>
<td>324 (13)</td>
<td>275 (11)</td>
<td>555 (21)</td>
</tr>
<tr>
<td><strong>Reread</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cont.-Match</td>
<td>1011 (73)</td>
<td>389 (29)</td>
<td>324 (26)</td>
<td>599 (39)</td>
</tr>
<tr>
<td>Cont.-Mismatch</td>
<td>1373 (164)</td>
<td>462 (45)</td>
<td>322 (29)</td>
<td>682 (44)</td>
</tr>
<tr>
<td><strong>Total Times</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cont.-Match</td>
<td>1447 (63)</td>
<td>483 (21)</td>
<td>363 (19)</td>
<td>834 (36)</td>
</tr>
<tr>
<td>Cont.-Mismatch</td>
<td>1772 (132)</td>
<td>532 (30)</td>
<td>400 (21)</td>
<td>931 (40)</td>
</tr>
</tbody>
</table>

Pre-critical Region

No effect of GENDERMATCH was observed in early measures (ts < 1). A significant effect of GENDERMATCH was observed in reread times, with the Control-Mismatch condition taking longer to read than the Control-Match condition ($\beta = 382.0$, s.e. = 173.0, t= 2.21). This pairwise effect achieved significance in total times, once again with Mismatch conditions taking longer to read than Match regions ($\beta = 332.0$, s.e. = 132.0, t= 2.52, p < .05).

Critical Pronoun Region

In the critical pronoun region no effects achieved significance (ts < 1.5).
Post-Pronoun Region

In the post-pronoun region no effects achieved significance (ts < 1.5).

Pronoun+2 Region

Two words after the pronoun a significant effect of GENDERMATCH was observed in total times ($\beta = 91.8$, s.e. = 43.9, $t = 2.09$), due to elevated RTs in Control-Mismatch conditions relative to Control-Match conditions. No other effects were found in other measures.

Discussion

Experiment 2 sought to determine whether grammatically inappropriate feature-matching QPs interfered with antecedent retrieval for a pronoun. The study contrasted the accessibility of a feature-matching referential NP that did not c-command the pronoun because it was embedded inside a relative clause with that of a QP in the same position. Conditions with corresponding non-matching 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. An offline acceptability study verified that comprehenders do not consider a feature-matching QP as a grammatical antecedent for a pronoun that the QP c-commands.

In the eye-tracking study, the first effects of note were in the test conditions. Across reading time measures, the condition in which there was a referential feature-matching antecedent was read
significantly more quickly at the critical pronoun region than in any other test conditions. A pairwise difference was observed between the Referential-Match condition and the Referential-Mismatch condition, in which there was no matching antecedent for the pronoun. No such effect was found between Quantificational conditions. When the antecedent was quantificational its gender concord with the pronoun did not have any effect on reading time measures. The facilitative effect of GenderMatch predicted by the interference account was not observed.

The second effects of note were in the Control conditions. Within the Control conditions, when the c-commanding QP matched the critical pronoun in gender, processing at and following the pronoun was facilitated in comparison to the condition in which the QP and pronoun mismatched in gender. This facilitation was observed in reread and total time measures, but not in earlier measures.

The pattern of effects observed in the test conditions show that comprehenders rapidly access referential feature-matching NPs embedded in a relative clause when retrieving antecedents for a pronoun not contained within the same relative clause. QPs in the same position were not accessed as potential antecedents for the pronoun, irrespective of morphological feature-match. This is consistent with the hypothesis that grammatically inappropriate feature-matching QPs do not interfere with antecedent retrieval. C-command information categorically restricts access to non-c-commanding QPs.

Given the absence of observable interference in the eye-tracking times, the slightly raised acceptability rating of the Quantificational-Match test condition relative to Quantificational-
Mismatch in Experiment 2a should not be interpreted as the result of interference. The rating more likely reflects eased coercion of a referent for the unheralded pronoun. The task of positing an extra-sentential referent for the unheralded pronoun in test sentences with a feature-matching QP is eased because the QP’s reference set provides salient and relevant properties that can be predicated of the extra-sentential reference. This minimizes cost of accommodating a new referent into the discourse (as represented by the sentence).

The differential time-course of mismatch effects between test and control conditions does raise questions regarding the accessibility of quantificational NPs generally. A number of possible explanations exist for the absence of a GENDERMATCH effect in early measures in the Control conditions, all of which will discussed at greater length in the general discussion. Two general kinds of explanations exist. First, it is possible that GENDERMATCH effects are masked in earlier measures by increases in processing time associated with successful binding. The second explanation would take late effects as indicative of a general delay, or dis-preference, for attempting bound variable readings for pronouns. This latter option seems unlikely given that successful binding in the previous experiment, and in other experiments (e.g. Cunnings et al. 2012), displays a time-course comparable to coreference resolution. Regardless of the interpretation of the late effects, however, it is clear that the feature content of structurally inappropriate quantificational NPs does not interfere with antecedent retrieval for pronouns.

**General Discussion**

**Summary of Results**

The main purpose of the present study was to assess whether non-c-commanding
morphologically feature-matching QPs are accessed during the retrieval of potential antecedents for a pronoun 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 antecedent retrieval was capable of distinguishing between structurally appropriate and structurally inappropriate QPs. Second, we asked if structurally inappropriate QPs interfered with antecedent retrieval in any capacity. Although the findings serve to enrich our general understanding of which grammatical constraints are and are not obeyed in real-time sentence comprehension, the experiments were primarily intended to probe 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.

Offline judgment studies corroborated the consensus from the formal linguistics literature that participants do not consider a binding relationship between a pronoun and a feature-matching QP that does not c-command that pronoun. In Experiments 1a and 2a, sentences containing a pronoun and a non-c-commanding feature-matching QP were rated as unacceptable as sentences that contained no feature-matching NP at all. In Experiment 1b participants reliably chose a referential reading for a pronoun that was in a structurally inappropriate position to be bound by the pronoun.

Experiment 1c showed that antecedent retrieval for a critical pronoun was capable of distinguishing structurally appropriate QPs that overtly c-commanded the pronoun from structurally inappropriate QPs that did not c-command the pronoun. Experiment 2b demonstrated that the advantage for structurally appropriate QPs reflected a categorical ban on accessing
grammatically inappropriate QPs. Non-c-commanding feature-matching QPs did not interfere with the retrieval of a pronoun’s antecedent.

*The Role of Relations in Retrieval*

The results suggest that retrieval respects the c-command constraint on bound-variable pronoun readings. Those QPs that the grammar excludes from consideration on the basis of their syntactic position in relation to a pronoun are not accessed by antecedent retrieval operations.

The results pose a *prima facie* challenge to cue-based retrieval mechanisms. If c-command relations are the relevant determinant of a QP’s grammatical accessibility and cue-based mechanisms are incapable of using relations as cues, it would appear that such models are unequipped to explain the retrieval sensitivity observed.

The experiments above could be interpreted as constituting preliminary support for the use of a different retrieval method in accessing c-commanding quantificational antecedents. The results could be modeling using a serial search procedure (Sternberg 1966; McElree & Dosher 1989, 1993; McElree, Foraker & Dyer 2003) 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 et al. 2013 for reflexive licensing). We note that although serial search mechanisms permit a straightforward implementation of a c-command constraint, the same mechanism is not well suited for modeling the retrieval of referential antecedents for pronouns because the distribution of such antecedents is not limited to the c-
command path. Any procedure that was restricted to the c-command path would fail to find non-c-commanding referential antecedents. Yet, Experiments 1 and 2 demonstrate that comprehenders display no difficulty in accessing non-c-commanding referential antecedents. If antecedent retrieval for c-commanding antecedents employs a serial search strategy, it would seem that a supplemental retrieval procedure would be required to identify non-c-commanding referential antecedents. A 'dual retrieval' strategy of this sort could have two serial search procedures run in parallel: one procedure that traversed the c-command domain to identify potential binders and a second procedure that searched the linear string for potential referential antecedents.

Serial search has a distinct temporal profile. Retrieval times are predicted to increase as a function of the distance between the retrieval probe and the retrieval target. If antecedent retrieval employed a serial search procedure, distance between a pronoun and its antecedent (either measured in number intervening structural nodes or in terms of simple linear distance) should influence speed of antecedent retrieval. This prediction is at least consistent with findings from previous studies that have shown distance effects for antecedent retrieval (Clark & Sengul 1979; Ehrlich 1983; Ehrlich & Rayner 1983; O'Brien 1987). The exact relevance of these distance effects in determining the mechanism of retrieval, however, is open to interpretation because the reading time measures used do not provide an independent index of retrieval speed. The effects observed could either represent faster retrieval speeds or facilitated integration for closer antecedents. Future work is required to test whether distance in fact influences antecedent retrieval speed.
To accommodate the results within the confines of a cue-based architecture would require a reconceptualization of the c-command constraint and adjustment to standard cue-combinatorics schemes. First, because c-command relations cannot be employed as retrieval cues, cue-based models must recode the relational constraint into a constraint on feature-match. Kush (2013) proposed that a feature, ACCESSIBLE, can be used to distinguish appropriate potential antecedents. ACCESSIBLE is a single feature assigned to all NPs (regardless of their quantificational/referential features) that are eligible retrieval targets relative to the current state of the parser. A structure-dependent dynamic update procedure can de-activate an NP's ACCESSIBLE feature at the point during incremental processing that the parser determines that the NP can no longer participate in later dependencies. QPs remain marked ACCESSIBLE so long as the parser is currently in their c-command domain. Upon leaving the c-command domain of a QP, the parser retrieves that QP and removes its accessibility feature, rendering it inert for all subsequent retrievals. Referential antecedents are never de-activated in this way, so they remain accessible regardless of the parser's current state.

The account of retrieval in terms of the feature ACCESSIBLE does not require encoding of relational or item-to-item information on individual items itself. Nor does it offer a feature-based translation of c-command per se. Reference to c-command resides in the ACCESSIBLE update function and not in the feature vocabulary. Importantly, this account does not require distinct retrieval mechanisms for referential and quantificational/binding antecedents. At any point, the cue-based retrieval mechanism is limited to retrieving only ACCESSIBLE NPs. Moreover, whereas a serial search mechanism can be thought of computing an antecedent's structural accessibility with respect to an individual pronoun 'on the fly' through its search procedure, the
ACCESSIBLE account effectively 'pre-computes' an antecedent's viability prior to ever encountering the pronoun.

Within the context of a cue-based retrieval architecture, the absence of partial-match interference from structurally inaccessible distractors suggests that syntactic cues can serve a gating function; NPs that otherwise match the retrieval probe in features are not retrieved if they mismatch the probe on syntactic features. 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). This equal-weight linear cue combinatorics scheme predicts a degree of partial match interference not observed in the foregoing studies. As discussed by Van Dyke and McElree (2011), there are two methods for modeling gating effects within a cue-based retrieval system: one can employ a weighted linear cue combinatorics scheme or a multiplicative scheme. Under a weighted linear scheme, the syntactic cue responsible for distinguishing appropriate from inappropriate antecedents could be disproportionately weighted relative to morphological cues. 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.

**Conclusion**

Our study asked two questions concerned with the ability of c-command information to guide pronoun antecedent retrieval. First, we tested whether antecedent retrieval is capable of distinguishing c-commanding QPs from non-c-commanding QPs. Second, we investigated
whether antecedent retrieval is subject to any measurable from intrusion effects from feature-matching non-c-commanding QPs. Results from two eye-tracking studies suggest that retrieval can distinguish structurally inappropriate from structurally appropriate QPs, and that retrieval is not subject to partial-match interference from structurally inappropriate QPs with matching morphological features.

The findings present a difficult, though not insuperable, challenge to modeling antecedent retrieval using traditional cue-based models of retrieval. Because relations such as c-command are difficult to use as retrieval cues in such architectures, an alternative method to make the grammatical distinction must be devised. Moreover, the results also seem to motivate a cue-combinatorics scheme that allows syntactic constraints to be more heavily weighted than other constraints in determining NP accessibility (as argued by Van Dyke and McElree 2011).

References


