

Variable binding and coreference in sentence comprehension: evidence from eye movements

Article

Published Version

Creative Commons: Attribution 3.0 (CC-BY)

Open Access

Cunnings, I., Patterson, C. and Felser, C. (2014) Variable binding and coreference in sentence comprehension: evidence from eye movements. *Journal of Memory and Language*, 71 (1). pp. 39-59. ISSN 0749-596X doi: <https://doi.org/10.1016/j.jml.2013.10.001> Available at <http://centaur.reading.ac.uk/35306/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1016/j.jml.2013.10.001>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



Variable binding and coreference in sentence comprehension: Evidence from eye movements



Ian Cunnings^{a,*}, Clare Patterson^b, Claudia Felser^b

^a Psychology and Clinical Language Sciences, University of Reading, United Kingdom

^b Potsdam Research Institute for Multilingualism, University of Potsdam, Germany

ARTICLE INFO

Article history:

Received 25 August 2012

revision received 16 October 2013

Available online 14 November 2013

Keywords:

Pronoun resolution

Eye movements

Reading

Memory retrieval

ABSTRACT

The hypothesis that pronouns can be resolved via either the syntax or the discourse representation has played an important role in linguistic accounts of pronoun interpretation (e.g. Grodzinsky & Reinhart, 1993). We report the results of an eye-movement monitoring study investigating the relative timing of syntactically-mediated variable binding and discourse-based coreference assignment during pronoun resolution. We examined whether ambiguous pronouns are preferentially resolved via either the variable binding or coreference route, and in particular tested the hypothesis that variable binding should always be computed before coreference assignment. Participants' eye movements were monitored while they read sentences containing a pronoun and two potential antecedents, a c-commanding quantified noun phrase and a non c-commanding proper name. Gender congruence between the pronoun and either of the two potential antecedents was manipulated as an experimental diagnostic for dependency formation. In two experiments, we found that participants' reading times were reliably longer when the linearly closest antecedent mismatched in gender with the pronoun. These findings fail to support the hypothesis that variable binding is computed before coreference assignment, and instead suggest that antecedent recency plays an important role in affecting the extent to which a variable binding antecedent is considered. We discuss these results in relation to models of memory retrieval during sentence comprehension, and interpret the antecedent recency preference as an example of forgetting over time.

Crown Copyright © 2013 Published by Elsevier Inc. All rights reserved.

Introduction

Successful real-time sentence and discourse comprehension requires linking pronouns to their likely antecedents as accurately as possible. In linguistic theory it has been posited that pronoun interpretation can be resolved in different ways (compare Bosch, 1983; Evans, 1980; Grodzinsky & Reinhart, 1993; Heim, 1992; Reinhart, 1983; Reuland, 2001, 2011, among others). While the precise details of these various theoretical accounts differ, a

core idea is that pronoun reference can be assigned in either the discourse representation, via *coreference* assignment, or via syntactically-mediated *variable binding*. To illustrate these two different types of dependency, consider first the examples in (1).

- (1a) The boy smiled. He was happy.
 (1b) Every boy smiled. He was happy.

In (1a), the pronoun *he* can refer back to *the boy* without any difficulty. By contrast, interpreting *he* in example (1b) as referring to the quantified phrase (QP) *every boy* is problematic, and the pronoun in this case is likely to be interpreted as referring to another (unmentioned) antecedent.

* Corresponding author. Address: School of Psychology and Clinical Language Sciences, University of Reading, Whiteknights Campus, Reading RG6 7BE, United Kingdom.

E-mail address: i.cunnings@reading.ac.uk (I. Cunnings).

Non-quantified noun phrases (NPs) such as *the boy* in (1a) are said to be referential in that they can pick out a specific individual in the discourse representation. When both an NP and a pronoun pick out the same individual from the discourse, they are said to corefer. The link between *he* and *the boy* in (1a) is thus said to be established via coreference assignment. Quantified phrases (QP) such as *every boy*, on the other hand, are usually assumed to be non-referential, in the sense that a QP does not single out a specific individual in the discourse. QPs cannot therefore enter into a coreference relationship with a pronoun, but the two can enter into an interpretive dependency through variable binding.

Unlike coreference relationships, variable binding is hypothesised to only hold intra-sententially. An important condition of variable binding is the requirement that the QP must c-command the pronoun. C-command describes a relationship between constituents in a structural representation of a sentence based on the notion of hierarchical dominance. In the standard definition, a constituent c-commands its sister constituents and any constituents that these dominate (Reinhart, 1983). In (1b) above, the c-command requirement of variable binding is not satisfied. In (2a,b) however, the pronoun is c-commanded by the QP, and as such can be bound by it in both cases.

- (2a) Every boy wished that he was happy.
 (2b) Every boy who knew John wished that he was happy.

The pronoun in (2b) is multiply ambiguous in that it can potentially be linked to the QP *every boy* via variable binding, to the proper name *John* through coreference assignment, or to another inferred discourse antecedent outside the current sentence. These two hypothesised routes to pronoun resolution also have interpretive consequences, as illustrated in ambiguous sentences such as (3).

- (3) During the clash, the footballer injured his leg and the referee did too.

Interpretation of (3) requires reconstruction of the elided verb phrase *injured his leg*, but this reconstruction is ambiguous as to whether the referee injured his own leg or whether the referee also injured the footballer's leg. The first interpretation (the referee injured his own leg) involves a copying of the *relationship* between the pronoun and its antecedent, and as such involves variable binding, whereas the second interpretation (the referee injured the footballer's leg) involves coreference between the reconstructed pronoun and a specific, referential antecedent and as such involves discourse-based coreference assignment.

While the binding/coreference distinction has been discussed extensively in the theoretical syntax and semantics literature, how it may affect online language processing has received far less attention. From a psycholinguistic perspective, one empirical issue relates to the question of which of these two types of relation (if any) is easier or quicker to compute during processing. Another, more general issue is

whether two qualitatively different memory search or retrieval mechanisms are required for each type of dependency, or whether both types of relation can be established via the same mechanism. The present study primarily addresses the first question, by examining whether readers prefer to initially link an ambiguous pronoun to either a variable binding or coreference antecedent during processing, but we will discuss the possible implications of our findings for models of memory search as well.

Resolving ambiguous pronouns

The question of whether the variable binding or coreference interpretation should be preferred in sentences such as (3) has been widely debated in the theoretical linguistics literature. Although the exact nature of each proposal differs, a number of researchers have claimed that syntactically mediated dependencies should be easier to compute, and hence favoured, over discourse-based ones (e.g. Avrutin, 1994, 1999; Burkhardt, 2005; Foley, Nunez del Prado, Barbier, & Lust, 2003; Grodzinsky & Reinhart, 1993; Vasić, Avrutin, & Ruigendijk, 2006). One influential account is Reuland's (2001, 2011) *Primitives of Binding* (POB) framework. Reuland claims that a preference for variable binding results from a relative economy hierarchy of referential dependencies, in which dependencies determined by purely syntactic processes (such as those involving 'standard' types of reflexive) are the easiest to compute, followed by those that require variable binding, with discourse-based coreference assignment being the most costly option. This economy principle is outlined in (4) below (adapted from Reuland, 2011: 338).

- (4) *Economy Principle*
 i. Minimise unresolved dependencies
 ii. Syntax < logical syntax < discourse

The economy principle thus predicts that the variable binding interpretation of sentences such as (3) should be preferred, and results from comprehension and reading time data have supported this hypothesis (Foley et al., 2003; Koornneef, Avrutin, Wijnen, & Reuland, 2011; Vasić et al., 2006; but see Frazier & Clifton, 2000 for mixed results).¹ A similar economy-based hierarchy can also be found in Optimality Theory (OT; see Prince & Smolensky, 2004) approaches to anaphora, where outputs are evaluated according to a set of soft (violable) constraints. For example, Hendriks and Spender's (2006) OT approach to binding incorporates a 'referential economy' constraint which states that reflexives are preferred to pronouns, which are in turn preferred to R-expressions (see also Burzio, 1998). Such an account also predicts that variable binding should be easier than coreference assignment.

¹ The assumption that binding and coreference relations are established at different levels of linguistic representation has been called into question by some however (e.g. Heim, 1998, 2007). Although it is not clear exactly what predictions can be derived from this proposal with regard to comprehenders' interpretive preferences or the time-course of pronoun resolution, we will briefly return to this point in the discussion of our findings.

Although not originally formulated as a model of online language processing, more recently it has been claimed that the POB economy principle provides not only an account of interpretive preferences of sentences such as (3), but also the time-course of pronoun resolution during language comprehension (Koornneef, 2008, 2010; Koornneef, Wijnen, & Reuland, 2006; Koornneef et al., 2011). Evidence from studies investigating the time-course of reflexive anaphor resolution suggests that reflexive binding relations, which under the POB approach are the easiest to compute, can indeed be established extremely quickly (e.g. Sturt, 2003; Xiang, Dillon, & Phillips, 2009). However, while reflexives have to be bound by a c-commanding antecedent in the local syntactic domain, non-reflexive pronouns are ambiguous and do not necessarily need to be syntactically bound.

Koornneef (2008) investigated whether variable binding is preferred over coreference assignment during pronoun resolution, in ambiguous cases when multiple antecedents are available in the discourse. Consider (5) below (from Koornneef, 2008: 136).

- (5) Een werkdag in de fabrieken langs het Noordzeekanaal is altijd erg zwaar.
(A working day in the factories along the North Sea Canal is always very tough.)

a. *Variable binding bias*

Voorals vandaag konden veel arbeiders, waaronder de oude Paul, het nauwelijks aan. Iedere arbeider die net als Paul bijna geen energie meer had, vond het heel erg fijn dat hij wat eerder naar huis mocht vanmiddag.
(Especially today a lot of workers, among them the old Paul, could barely cope. Every worker who just like Paul was running out of energy, thought that it was very nice that he could go home early this afternoon.)

b. *Coreference bias*

Voorals vandaag kon de oude arbeider Paul het nauwelijks aan. Iedere arbeider die zag dat Paul bijna geen energie meer had, vond het heel erg fijn dat hij wat eerder naar huis mocht vanmiddag.
(Especially today the old worker Paul could barely cope. Every worker who noticed that Paul was running out of energy, thought that it was very nice that he could go home early this afternoon.)

In both (5a,b), the pronoun *he* is ambiguous, and can be linked to either the QP *every worker*, which c-commands the pronoun, or the definite antecedent *Paul*, which does not c-command the pronoun and thus can only be linked to it via coreference. In (5a), the context biases towards the variable binding antecedent *every worker*, while (5b) is biased towards the coreference antecedent *Paul*. Koornneef hypothesised that if variable binding is always initially preferred over coreference assignment, reading times at or just after the pronoun in (5b) should be longer than in (5a). Although the coreference interpretation is ultimately favoured in (5b), Koornneef predicts that the pronoun should

initially be linked to a potential variable binder, i.e. the c-commanding QP, irrespective of the discourse context. In an eye-movement study in Dutch using materials as in (5), Koornneef observed longer second-pass reading times at the pronoun region (*that he*) in (5b) than (5a), a finding that was taken to support the hypothesis that variable binding relations are computed before coreference assignment.

Although these data are compatible with the predictions of the POB approach, other interpretations that do not rely on the specific hypothesised time-course of events are also possible. For example, the shorter reading times in (5a), in which the context favoured the QP, could be a result of a first-mention or subjecthood advantage (Järvikivi, van Gompel, Hyönä, & Bertram, 2005), or alternatively a preference for pronouns referring to main clause rather than subordinate clause antecedents (Cooreman & Sanford, 1996). The two critical sentences in (5a,b) also differ in their syntactic complexity, with (5b) involving an additional level of clausal embedding, which may have added to the relative difficulty of retrieving the (contextually favoured) coreference antecedent in (5b). Finally, in (5b) both antecedents appear as subjects, while in (5a) only the QP is a subject. The increased reading times in (5b) might thus reflect competition between antecedents when both are subjects. Due to the nature of the experimental design, it is difficult to choose between these alternative explanations, and it is difficult to tell precisely when during processing each potential antecedent was considered.

There is also existing evidence that potentially conflicts with the prediction that variable binding relations should be less costly than those involving coreference. For example, Carminati, Frazier, and Rayner (2002) showed that, all other things being equal, linking a pronoun to a QP antecedent incurs longer reading times than linking a pronoun to a non-quantificational definite noun phrase (see also Burkhardt, 2005). Some of the findings reported by Koornneef (2008) also support this observation. In addition to the ambiguous conditions in (5), Koornneef's study also included two unambiguous conditions in which there was only one antecedent for the pronoun; either a QP, as in (6a), or a proper name, as in (6b) (Koornneef, 2008: 137).

- (6a) Iedere arbeider die bijna geen energie meer had, vond het heel erg fijn dat hij wat eerder naar huis mocht vanmiddag.
(Every worker who was running out of energy, thought that it was very nice that he could go home early this afternoon.)
- (6b) Paul had bijna geen energie meer. Het was heel erg fijn dat hij wat eerder naar huis mocht vanmiddag.
(Paul was running out of energy. It was very nice that he could go home early this afternoon.)

In contrast to the predictions of the POB's economy principle that variable binding relations should be less costly than coreference assignment, the results of these unambiguous sentences indicated longer second-pass times at a spillover region following the pronoun for the condition containing a QP antecedent rather than a proper name.

There are thus conflicting results with regards to preferences for either variable binding or coreference assignment. While Koornneef's (2008) study suggests that when multiple antecedents are available variable binders may initially be preferred, other interpretations of these data are possible, and other studies have shown that linking pronouns to QPs is more costly than linking them to non-quantified antecedents (Burkhardt, 2005; Carminati et al., 2002).

Memory retrieval during sentence comprehension

That pronoun interpretation can be resolved in different ways poses a challenge for the memory retrieval processes involved in accessing the antecedent of a pronoun during sentence comprehension. While retrieval of a coreference antecedent is likely to rely on a set of discourse-based retrieval cues defined in terms of discourse prominence, or as a result of matching (person, number, or gender) features, accessing a variable binder relies on syntactic cues based on structural relations between constituents. While a variety of different factors are known to influence pronoun resolution, how ambiguity between potential variable binding and coreference antecedents is resolved during processing has to date received little attention. The hypothesis that accessing a variable binder should take temporal precedence over a coreference antecedent (Koornneef, 2008) can be seen as one potential solution to this ambiguity, in that the decision about whether a variable binding or coreference antecedent is preferentially retrieved is based upon the position of the antecedent in the economy hierarchy illustrated in (4).

Research on memory retrieval has outlined two qualitatively different search mechanisms. In serial searches, memory representations are sequentially searched in a step-by-step fashion until the appropriate information is retrieved. Such searches can be contrasted with another type of retrieval mechanism in which representations are accessed directly in a content-addressable memory (CAM) architecture, where memory retrieval occurs when the content of a representation in memory is matched with those of a set of retrieval cues (see McElree, 2006, for an overview of serial search vs. content-addressability). One key property of serial search is that retrieval speed is dependent on the position of an item in the search path, such that items that appear later in the search stack are more slowly retrieved. On the other hand, in a CAM architecture content-matching representations are accessed immediately, but retrieval can be impeded by similarity-based interference when multiple representations in memory match the retrieval cues (see e.g. Lewis, Vasishth, & Van Dyke, 2006; McElree, 2000; McElree, Foraker, & Dyer, 2003; Van Dyke & McElree, 2006, 2011).

It has recently been claimed that memory retrieval during sentence comprehension is mediated by a direct-access CAM architecture (see Lewis et al., 2006, for an overview). Support from this view has come from evidence of similarity-based interference (Van Dyke, 2007; Van Dyke & McElree, 2006) and length-invariant access times for memory retrieval during sentence processing (Martin & McElree,

2008; McElree, 2000; McElree et al., 2003). For example, McElree et al. (2003) used a speed-accuracy tradeoff paradigm and found that dependency length in filler-gap dependencies did not affect the speed of retrieval, but it did influence the accuracy with which the filler was retrieved (McElree et al., 2003). This decrease in retrieval accuracy as dependency length increases suggests that the fidelity of a representation retrieved during language processing degrades over time as a result of forgetting (see Van Dyke & Johns, 2012, for review). Dependencies that have been argued to be mediated via content-addressable searches include subject-verb agreement (Wagers, Lau, & Phillips, 2009), filler-gap dependencies (McElree et al., 2003) and verb-phrase ellipsis (Martin & McElree, 2008), as well as pronoun resolution (Foraker & McElree, 2007).

Foraker and McElree (2007) investigated how prominence influences antecedent retrieval. In particular, they compared whether focused antecedents are held in a distinct memory state in the focus of attention, vs. being encoded with a particularly strong representation. In a speed-accuracy tradeoff paradigm, they observed that focused antecedents were more accurately retrieved than non-focused antecedents, but that retrieval speed did not differ between the two. Foraker and McElree thus concluded that focused antecedents are not kept in a privileged state in the focus of attention, but rather are encoded with more distinct representations in memory that increase the likelihood, but not speed, of them being retrieved.

The hypothesis that accessing a variable binder should take temporal preference over a coreference antecedent during pronoun resolution (Koornneef, 2008; Reuland, 2011) can potentially be taken as an example of a memory retrieval mechanism that relies on a serial search, as the priority in which antecedents are retrieved is dependent upon their relative position in the search path.

Note however that antecedents which c-command a given pronoun are often in particularly prominent positions (e.g. matrix subject of a sentence). Following the results of Foraker and McElree (2007), the fact that c-commanding antecedents might be particularly discourse prominent leads to the possibility that a c-commanding variable binder may be retrieved preferentially over a non c-commanding coreference antecedent because it has a more distinct representation in memory, rather than as a result of its syntactic position. In this way, prominence could potentially act as a proxy for the c-command requirement of variable binding in a content-addressable architecture. The role of prominence in anaphora resolution is explicitly articulated in discourse prominence theory (DPT) (Gordon & Hendrick, 1998), which claims that pronouns trigger retrieval of the most prominent antecedent within a piece of discourse. In principle, DPT could be implemented as a serial search which checks antecedents serially in terms of prominence, or in a CAM architecture in which more prominent antecedents are encoded with more distinct representations in memory. Either way, prominence in DPT is syntactically defined and is related to the height of an antecedent in the syntactic tree, with antecedents higher in the syntactic tree being more prominent. Although

syntactic prominence in DPT is closely related to the notion of c-command, unlike the POB approach which claims that c-commanding variable binders are *always* initially preferred, other factors, such as whether or not an antecedent appears inside another noun phrase, can also affect prominence in DPT.

DPT claims that recency is another factor that may influence prominence. Streb, Henninghausen, and Rösler (2004) provide evidence from event-related potentials indicating that linking a pronoun to a coreference antecedent across one or more intervening sentences becomes progressively more difficult. This antecedent recency effect can be attributed to greater difficulty in retrieving a more degraded representation from memory as the length of the interpretive dependency increased. That representations retrieved during language comprehension can degrade over time might also have consequences for cases when multiple possible representations in memory are grammatically accessible for retrieval, as in the case of an ambiguous pronoun with multiple potential antecedents. Other things being equal, forgetting may increase the probability of a more recent antecedent being retrieved.

Against this background, we carried out two reading experiments using gender congruence as an index of dependency formation (see e.g. Badecker & Straub, 2002; Cunnings & Felser, 2013; Kazanina, Lau, Liberman, Yoshida, & Phillips, 2007; Sturt, 2003; Van Gompel & Liversedge, 2003). In both eye-movement experiments, we manipulated gender congruence between a pronoun and two potential antecedents; a variable binder that c-commanded the pronoun, and a coreference antecedent that did not. This design allows us to more directly assess the relative time-course in which each antecedent is retrieved than was possible in the design used by Koornneef (2008). In Experiment 1, the variable binding antecedent occurred linearly more distant to the pronoun than the coreference antecedent, whilst in Experiment 2 the linear order of antecedents was reversed. The two online experiments were also complemented by an offline antecedent choice task to examine comprehenders' ultimate interpretive preferences. Our primary aim was to investigate whether linking a pronoun to one particular type of antecedent, either a variable binder or coreference antecedent, is initially preferred.

Experiment 1

To examine the time-course of pronoun resolution, we monitored participants' eye-movements as they read a series of texts as shown in (7) below. A design similar to that used by Sturt (2003) and Cunnings and Felser (2013) was adopted in which gender congruence between the pronoun and each antecedent was manipulated in a 2×2 design yielding four conditions.

(7a) *QP Match, Name Match*

The squadron paraded through town. Every soldier who knew that James was watching was convinced that he should wave as the parade passed. The entire town was extremely proud that day.

(7b) *QP Match, Name Mismatch*

The squadron paraded through town. Every soldier who knew that Helen was watching was convinced that he should wave as the parade passed. The entire town was extremely proud that day.

(7c) *QP Mismatch, Name Match*

The squadron paraded through town. Every soldier who knew that Helen was watching was convinced that she should wave as the parade passed. The entire town was extremely proud that day.

(7d) *QP Mismatch, Name Mismatch*

The squadron paraded through town. Every soldier who knew that James was watching was convinced that she should wave as the parade passed. The entire town was extremely proud that day.

Gender congruence between the pronoun and each antecedent was manipulated such that in (7a,b) the QP (*every soldier*) matched in stereotypical gender with the pronoun whereas in (7c,d) it did not. Similarly, in (7a,c) the proper name (*James/Helen*) antecedent additionally matches the gender of the pronoun whereas in (7b,d) there is a gender mismatch.

If, as predicted by the POB approach (Koornneef, 2008), variable binders are accessed before coreference antecedents, comprehenders should initially attempt to link the pronoun to the QP only. If this is the case, reading times during or shortly after the initial inspection of the pronoun should be longer in (7c,d), where there is a stereotypical gender mismatch between the pronoun and QP, in comparison to (7a,b). Evidence of attempted coreference assignment between the pronoun and proper name should be either absent or in comparison delayed to effects of the variable binder, such that influence of the gender of the proper name antecedent, as indexed by comparing reading times in (7a,c) to (7b,d), should be restricted to comparatively later reading time measures, such as second-pass measures, or to regions of text downstream of the pronoun.

A similar set of predictions might also be derived from DPT (Gordon & Hendrick, 1998). In (7), although both the QP and proper name are subjects, the QP is the matrix subject of the critical sentence and is higher in the syntactic tree than the proper name, which appears inside a noun phrase modifier. These factors may make the QP more prominent which in turn will favour its initial retrieval over the proper name antecedent. Note however, that the proper name antecedent is linearly closer to the pronoun, which may act to favour it over the QP.

An alternative set of predictions can be made if agreement acts as a highly ranked cue to retrieval (Wagers et al., 2009). In this case, upon encountering a pronoun a search will be initiated for antecedents that match in gender and (grammatical) number. In this case, we might expect only (7d) to cause processing difficulty, as in all other conditions a gender and number matching antecedent can be found, by linking the pronoun to either the QP

or proper name in (7a), the QP in (7b) and the proper name in (7c). Competition between both antecedents might also potentially occur in condition (7a), when both antecedents match the pronoun's gender. This could in turn lead to longer reading times in multiple match condition (7a), in comparison to conditions (7b) and (7c), when only one antecedent matches the gender of the pronoun.²

However, in addition to agreement cues, semantic cues to pronoun resolution are also likely to guide memory retrieval. In this regard, the variable binding QP antecedent, which is conceptually plural but lacks a unique real-world referent, provides a less obvious match semantically for the singular pronoun *s/he* compared to the (semantically singular) coreference antecedent, the proper name. Recall that existing evidence has shown that linking a pronoun to a non-quantified noun phrase is less costly than linking it to a QP antecedent (Burkhardt, 2005; Carminati et al., 2002). These results may suggest that antecedents, such as proper names, that match the semantic properties of singular pronouns more closely than QPs may be preferred. In this case, it might be that memory retrieval at the pronoun initially favours the proper name, which matches the pronoun in terms of both grammatical and semantic number. Additionally, definite proper name antecedents also match the definiteness properties of pronouns more closely than QPs. If definiteness and number are more highly weighted retrieval cues than gender, it could be that the proper name antecedent is preferentially retrieved, irrespective of its gender. If this is indeed the case, this should lead to longer reading times in conditions (7b,d), when the proper name mismatches in gender with the pronoun, in comparison to (7a,c), when there is a gender match, as an index of processing difficulty resulting from the detection of gender incongruity once the antecedent has been retrieved. That the proper name might be preferentially retrieved might also be expected if antecedent recency affects pronoun resolution.

Method

Participants

Twenty-seven native English speakers (nine males, mean age 21) were paid a small fee to participate in the experiment. All participants had normal or corrected to normal vision and were recruited from the University of Essex community.

Materials

Twenty-four sets of experimental items were constructed as in (7). Gender congruence between the proper name and pronoun was manipulated using male and female proper names, whereas a stereotypical gender

manipulation was used between the QP and pronoun, using only highly gender biased nouns based on ratings from a previous norming study (Cunnings & Felser, 2013). The full set of experimental items is provided in the Appendix.

In addition to the experimental items, 60 filler texts were also constructed. These included distractor items that contained different types of pronouns in different structural configurations to those in the main experimental stimuli.

Procedure

The experimental and filler items were pseudo-randomised such that no two experimental items appeared adjacent to each other and were spread across four presentation lists in a Latin-square design. The experiment was divided into four blocks at which point participants could take a break if required. Forward and reverse orders within each block were constructed and the ordering of each block was different for each participant. The experiment began with five practice items to familiarise participants with the procedure. All items were presented in Courier New font, and displayed across up to three lines of text onscreen.

Eye movements were recorded using the EYELINK II system, which records participant eye movements via two cameras mounted on a headband at 500 Hz. Participant head movements are automatically compensated for by a third camera mounted in the centre of the headband which tracks the position of four LEDs on each corner of the computer screen. While viewing was binocular, the eye-movement data was recorded from the right eye only.

Each experimental session began with calibration of the eye-tracker on a nine-point grid. Before each trial, calibration was checked via a drift correction marker positioned above the first word of the next trial to be displayed. Participants were instructed to fixate upon this marker, and press a button to view the next trial. Any drift in the headset was automatically compensated for before presentation of the trial.

Participants read each text silently at their normal reading rate, pressing a button on a control pad once completed. Content questions requiring a yes–no push button response followed two thirds of all trials to ensure that participants paid attention to the content of the sentences, half of which required a 'yes' and half a 'no' response. The entire experiment lasted approximately 30–45 min in total.

Data analysis

The eye-movement record provides a rich source of information about the time-course of language processing at any point in a sentence (Rayner, 1998). To examine the time-course of pronoun resolution, we calculated reading times for four regions of text. As the pronouns used were all very short, the critical *pronoun region* consisted of the pronoun plus the preceding word, which was always the complementiser *that* (i.e. *that s/he*). We extended the pronoun region to the left rather than to the right as, given that the perceptual span of readers in

² Note that in Lewis et al.'s (2006) computational model of content-addressable search, memory retrieval is a 'winner-takes-all' process, in which only the most highly activated antecedent that matches the retrieval cues is retrieved. In this sense, although multiple antecedents may receive some degree of activation in (7a), only one antecedent, that which has the most activation at the point of retrieval, will be retrieved. If this is the case, the potential for 'competition' may not in itself lead to any measurable differences in reading times (see Dillon, 2011, for further discussion).

English is approximately eight characters (Rayner, 1998), fixations on the complementiser are likely to also include foveal processing of the pronoun. Extending the pronoun region to the right may have risked mixing first-pass processing of the pronoun with spillover effects at the following word. A *spillover region* consisted of the two words following the pronoun (e.g. *should wave* in (7)), the *prefinal region* consisted of the next two following words (*as the*) and the *final region* the rest of the critical sentence (*parade passed*).

Four reading time measures are reported for each region. *First-pass reading time* is the summed duration of fixations within a region during its first inspection, until it is exited to the left or right. *Regression path duration* is calculated by summing the duration of each fixation, starting with the first fixation when a region is entered from the left, up until but not including the first fixation in a region to the right. In addition to these two *first-pass* processing measures, we also calculated *rereading time*, a *second-pass* measure that includes all fixations within a region *after* it has been exited following the first-pass. *Total viewing time*, the summed durations of all fixations within a region, was also calculated as an overall measure of a region's processing load.

All trials in which track loss occurred were discarded, and regions which were initially skipped during reading were treated as missing data. For rereading time, trials in which a region was not refixated after the first-pass contributed a rereading time of zero to the calculation of averages. Prior to the calculation of these measures an automatic procedure merged short fixations of 80 ms or below that were within one degree of visual arc of another fixation. All other fixations of 80 ms or below, as well as those above 800 ms, were removed before further analysis.

Results

Overall accuracy to the comprehension questions was 92%, indicating that participants paid attention to the content of the sentences. Track loss accounted for less than 0.5% of the data, and skipping rates for the pronoun, spillover, prefinal and final regions were 20.4%, 6.5%, 11.0% and 6.3% respectively.

The eye-movement data were analysed using linear mixed-effects modelling (see Baayen, 2008; Baayen, Davidson, & Bates, 2008, for discussion) with the lme4 package in R. For each reading time measure, a model was fitted containing centred, fixed main effects of QP (match vs. mismatch), proper name (match vs. mismatch) and the QP * Proper Name interaction. In each case, the 'maximal' random effects structure, containing random intercepts for both subjects and items and by-subject and by-item random slopes for each fixed effect, was fitted (Barr, Levy, Scheepers, & Tily, 2013). If this maximal model failed to converge, the random slope parameter that accounted for the least amount of variance was removed and the model refitted until convergence was achieved. For each reading time measure, *p* values for each fixed effect were calculated using an upper bound of the *t* statistic (Baayen, 2008: 248).

Summaries of the reading time data and statistical analysis for Experiment 1 are presented in Tables 1 and 2 respectively.³

While there were no reliable effects observed in the first-pass measures at the pronoun region, a reliable main effect of proper name was observed in both rereading and total viewing times. Rereading times were significantly longer in conditions (7b,d), when the pronoun mismatched in gender with the proper name antecedent, in comparison to (7a,c), when there was a gender match (307 ms vs. 188 ms). The same pattern of results, with longer reading times for conditions (7b,d) in comparison to (7a,c), was also found in total viewing times (662 ms vs. 504 ms). There was no hint of any influence of the gender of the QP antecedent in any reading time measure at the pronoun region. This pattern of results is shown in Fig. 1(a), which shows total viewing times for the pronoun region.

A similar pattern of results was observed at the spillover region. Reading times were reliably longer in proper name gender mismatching conditions (7b,d) in comparison to gender matching conditions (7a,c) in regression path times (606 ms vs. 485 ms), rereading times (384 ms vs. 265 ms) and total viewing times (723 ms vs. 574 ms). No reliable influence of the gender of the QP was observed in any of these measures.

At the prefinal region, there was again a marginally significant main effect of proper name in the regression path times, with reading times tending to be longer following a gender mismatch between the pronoun and proper name antecedent in comparison to when there was a gender match (606 ms vs. 475 ms). At the final region, there was again a marginal trend of a main effect of proper name in regression path times, with longer reading times in conditions (7b,d) in comparison to (7a,c) (1170 ms vs. 1017 ms). No reliable influence of the gender of the QP antecedent was observed at either the prefinal or final region in any reading time measure.

Discussion

The results of Experiment 1 indicate clearly that readers preferred to link the pronoun to the proper name antecedent rather than the QP. In a number of reading time measures at various regions of text, we observed longer reading times when the proper name mismatched in gender with the pronoun in comparison to when there was a gender match. This effect was observed both at the critical pronoun region and at regions of text downstream of the pronoun. We found no reliable evidence of the QP ever being considered as a potential antecedent for the pronoun in any reported measure, at any region of text.

Our finding that participants initially attempted to link the pronoun to a proper name antecedent that did not c-command it rather than a c-commanding QP fails to sup-

³ To ensure that there were no reading time differences between conditions prior to the critical region, we also calculated first pass and regression path times for a *precritical region*, containing the text in between the proper name antecedent and critical pronoun. These results revealed no reliable effects for either measure (all *t* < 1.47, all *p* > .142), indicating that the differences observed at the critical region are unlikely to be a result of differences between conditions before the pronoun was encountered.

Table 1

Reading times for four eye-movement measures at four regions of texts in Experiment 1 (SDs in parentheses).

	First pass reading time	Regression path time	Rereading time	Total viewing time
<i>Pronoun region</i>				
QP Match, Name Match	304 (177)	420 (522)	195 (374)	499 (421)
QP Match, Name Mismatch	358 (260)	408 (318)	275 (421)	633 (537)
QP Mismatch, Name Match	329 (221)	391 (286)	180 (352)	508 (443)
QP Mismatch, Name Mismatch	352 (221)	437 (373)	339 (572)	691 (645)
<i>Spillover region</i>				
QP Match, Name Match	290 (175)	520 (803)	254 (395)	544 (444)
QP Match, Name Mismatch	332 (232)	606 (607)	363 (492)	695 (548)
QP Mismatch, Name Match	328 (221)	449 (343)	275 (510)	603 (564)
QP Mismatch, Name Mismatch	346 (293)	606 (584)	405 (641)	751 (688)
<i>Prefinal region</i>				
QP Match, Name Match	310 (176)	473 (561)	206 (284)	517 (332)
QP Match, Name Mismatch	316 (196)	599 (731)	261 (438)	577 (484)
QP Mismatch, Name Match	308 (175)	476 (515)	211 (375)	518 (404)
QP Mismatch, Name Mismatch	308 (179)	612 (843)	258 (391)	567 (450)
<i>Final region</i>				
QP Match, Name Match	367 (252)	869 (1002)	145 (276)	512 (399)
QP Match, Name Mismatch	366 (226)	1188 (2282)	202 (452)	568 (507)
QP Mismatch, Name Match	387 (249)	1165 (1807)	185 (349)	572 (416)
QP Mismatch, Name Mismatch	391 (277)	1151 (2687)	222 (390)	613 (503)

QP = Quantifier Phrase, Name = Proper Name.

Table 2

Summary of statistical analyses for four eye-movement measures at four regions of texts in Experiment 1.

	First pass reading time		Regression path time		Rereading time		Total viewing time	
	Estimate	<i>t</i>	Estimate	<i>t</i>	Estimate	<i>t</i>	Estimate	<i>t</i>
<i>Pronoun region</i>								
QP	7 (17)	0.42	7 (41)	0.18	23 (42)	0.54	23 (43)	0.54
Name	28 (21)	1.30	18 (43)	0.43	110 (36)	3.04**	140 (40)	3.51***
QP* Name	7 (39)	0.17	68 (81)	0.84	89 (89)	1.00	74 (88)	0.84
<i>Spillover region</i>								
QP	27 (24)	1.11	34 (60)	0.57	30 (39)	0.77	59 (43)	1.37
Name	32 (21)	1.57	119 (53)	2.32**	107 (44)	2.43**	136 (44)	3.13**
QP* Name	31 (41)	0.75	56 (109)	0.12	9 (72)	0.12	21 (81)	0.26
<i>Prefinal region</i>								
QP	3 (14)	0.24	7 (58)	0.12	1 (30)	0.04	1 (31)	0.04
Name	2 (14)	0.15	119 (66)	1.78*	50 (31)	1.63	51 (32)	1.62
QP* Name	1 (37)	0.04	19 (105)	0.18	15 (67)	0.22	8 (74)	0.12
<i>Final region</i>								
QP	19 (19)	1.00	290 (209)	1.39	29 (27)	1.06	46 (31)	1.49
Name	2 (21)	0.11	320 (186)	1.72*	45 (37)	1.21	47 (42)	1.12
QP* Name	4 (47)	0.09	30 (334)	0.09	12 (65)	0.18	8 (57)	0.15

QP = Quantifier Phrase, Name = Proper Name. Estimate = Model Estimate (SE in brackets).

* $p < .10$.** $p < .05$.*** $p < .001$.

port the hypothesis that variable binding relations are computed before coreference assignment. Our finding that gender mismatching proper name antecedents led to processing difficulty, even in condition (7c) when an alternative, stereotypical gender matching antecedent was available, suggests that the preference for a definite, grammatically and semantically singular antecedent may, at least initially, outweigh any preference for satisfying the pronoun's gender agreement cues. We also found no evidence that having multiple gender matching antecedents in the discourse led to any measurable processing difficulty or competition.

Several previous studies have reported an advantage for pronouns that refer to the first-mentioned referent in a piece of discourse (Arnold, Brown-Schmidt, & Trueswell, 2007; Gernsbacher, 1990; Gernsbacher & Hargreaves, 1988; Gordon, Grosz, & Gilliom, 1993). In a language such as English which has a predominant SVO word order, first-mention is often confounded with subjecthood. Järviö et al. (2005) showed that in Finnish, a language with freer word order where both SVO and OVS structures are possible, the antecedent search is guided both by subjecthood and by the order-of-mention of antecedents. In the current study, both potential antecedents grammatically func-

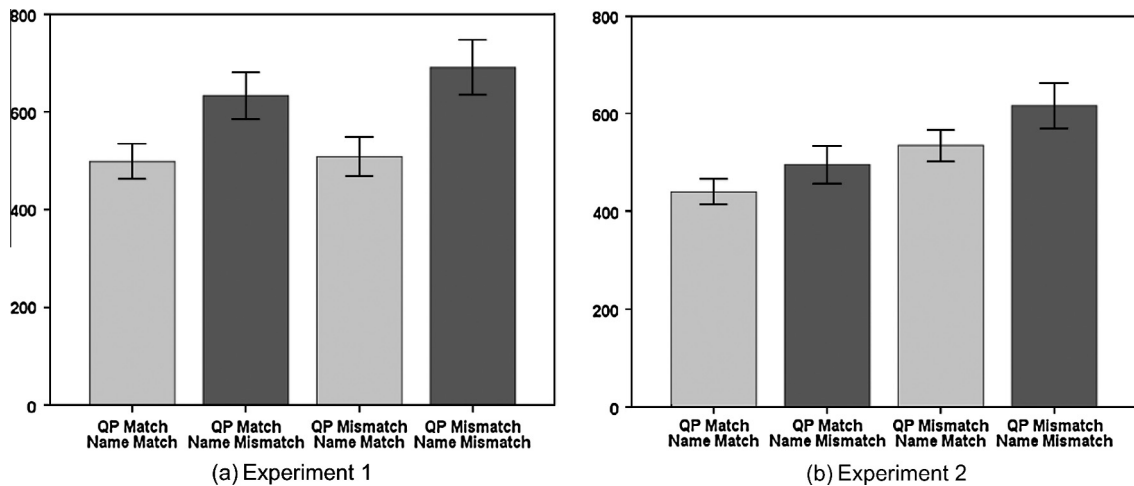


Fig. 1. Total viewing times (in ms) with standard errors for the pronoun region in Experiment 1 (a) and Experiment 2 (b), QP = Quantifier Phrase, Name = Proper Name.

tioned as subjects, and although the QP was the highest antecedent in the syntactic structure, the results from Experiment 1 provide no evidence that the first mentioned antecedent (the QP) was favoured at any point during processing. Instead, it is possible that the QP's relative distance from the pronoun played a more important role than first-mention here, or that quantified antecedents are generally dispreferred. Another alternative possibility could be that the preference observed for the proper name antecedent was a result of subtle properties of the discourse rather than antecedent recency. We come back to this point in the General Discussion. Either way, the results of Experiment 1 provide no support for the hypothesis that variable binding should always initially be preferred.

Experiment 2 was designed to further examine how antecedent recency interacts with variable binding and coreference assignment. If pronoun resolution is influenced by antecedent recency (Streb et al., 2004), we would expect to find a different pattern of results in Experiment 2, where the order of antecedents is reversed such that the QP is now linearly closer to the pronoun. Thus, if antecedent recency affects whether the QP is retrieved, then reading times in Experiment 2 should be affected by the stereotypical gender of the QP rather than the proper name antecedent. In contrast, if matching of grammatical and semantic number cues outweighs effects of recency and other retrieval cues such as gender agreement, then readers may again prefer to link the pronoun to the coreference antecedent despite the fact that this is now further away from the pronoun. This is also what might be expected if linking a pronoun to a non-pronominal QP antecedent is inherently more costly than linking it to a non-quantified antecedent (Burkhardt, 2005; Carminati et al., 2002), as readers may avoid linking pronouns to QPs whenever possible.

Experiment 2

The 24 texts in four conditions from Experiment 1 were adapted for Experiment 2. The linear order of antecedents

was reversed such that the QP antecedent was now linearly closer to the pronoun than the coreference antecedent, as illustrated by (8a–d).

(8a) QP Match, Name Match

The squadron paraded through town. It looked to James that every soldier was completely convinced that he should wave as the parade passed. The entire town was extremely proud that day.

(8b) QP Match, Name Mismatch

The squadron paraded through town. It looked to Helen that every soldier was completely convinced that he should wave as the parade passed. The entire town was extremely proud that day.

(8c) QP Mismatch, Name Match

The squadron paraded through town. It looked to Helen that every soldier was completely convinced that she should wave as the parade passed. The entire town was extremely proud that day.

(8d) QP Mismatch, Name Mismatch

The squadron paraded through town. It looked to James that every soldier was completely convinced that she should wave as the parade passed. The entire town was extremely proud that day.

Although the linear order of antecedents has been reversed, the QP antecedent is again in a structural position such that it c-commands the pronoun, while the proper name does not.⁴ The POB economy principle (3) again predicts that the QP antecedent should be considered before the coreference

⁴ A reviewer points out that although the proper name does not c-command the pronoun under traditional assumptions, some analyses would claim that it does, such that the proper name antecedent can also act as a variable binder in the constructions tested in Experiment 2 (compare e.g. Collins, 2005).

antecedent. In this case, we would expect to find longer reading times in condition (8c,d), when the pronoun mismatches in stereotypical gender with the QP antecedent, in comparison to conditions (8a,b), when there is a stereotypical gender match, during or shortly after the initial inspection of the pronoun region. DPT (Gordon & Hendrick, 1998) may also favour the QP, on the assumption that it is more salient than the proper name antecedent due to it being a finite clause subject. Recency would also favour the QP, as in Experiment 2 it is now linearly closer to the pronoun than the proper name antecedent.

Conversely, if proper names are generally preferred over quantified antecedents, either because they provide a better match for the definite properties of pronouns, or because linking a pronoun to a QP is more costly than linking it a proper name, then reading times should be longer when the proper name mismatches in gender with the pronoun in conditions (8b,d), in comparison to when there is a gender match in (8a,c).

Method

Participants

Thirty-one native English speakers (nine males, mean age 24) from the University of Essex community, none of whom took part in Experiment 1, were paid a small fee to take part in Experiment 2.

Materials

The 24 experimental items from Experiment 1 were adapted as in (8) (see Appendix). Sixty filler texts and five practice items were again included.

Procedure and data analysis

The order of items was pseudo-randomised, presented in four blocks and divided across four presentation lists as in Experiment 1. All other aspects of the procedure and data analysis in this experiment were the same as in Experiment 1.

Results

Overall accuracy to the comprehension questions was 91%. Track loss occurred in 1.6% of the data, and skipping rates of the pronoun, spillover, prefinal and final regions were 20.7%, 7.1%, 12.2% and 7.3% respectively. Summaries of the reading time data and statistical analysis for Experiment 3 are provided in Tables 3 and 4 respectively.⁵

Analysis of the first-pass reading times at the pronoun region revealed a marginally significant main effect of QP, with reading times tending to be longer in conditions (8c,d), when the QP mismatched in stereotypical gender with the pronoun, in comparison to conditions (8a,b), when there was a stereotypical gender match (332 ms vs. 294 ms). There was also an additional trend for longer first-pass reading times when the proper name antecedent mismatched in gender with the pronoun in comparison to when there was a gender match (319 ms vs. 297 ms), but

this was marginally significant. Measures including second-pass processing at the pronoun region revealed main effects of the QP only, marginal in rereading time and significant in total viewing time, with reading times again being longer in stereotypical gender mismatching conditions (8c,d) in comparison to stereotypical gender matching conditions (8a,b) (254 ms vs. 175 ms and 576 ms vs. 468 ms in rereading and total viewing times respectively). The total viewing time data from the pronoun region for this experiment are shown in Fig. 1(b).

While no reliable effects were observed in the first-pass measures at the spillover region, there was a significant main effect of QP in rereading times. Here, reading times were again longer in stereotypical gender mismatching conditions (8c,d) in comparison to mismatching conditions (8a,b) (282 ms vs. 215 ms). There was also a trend for longer rereading times when the proper name antecedent mismatched in gender with the pronoun (274 ms vs. 223 ms), but the main effect of proper name was only marginally significant. In total viewing time, there was again a marginal main effect of QP, with reading times being longer following a stereotypical gender mismatch in comparison to when there was a stereotypical gender match (574 ms vs. 529 ms).

No significant main effects or interactions were observed for any measure at the prefinal region, but at the final region regression path times we again observed a significant main effect of QP, with longer reading times in conditions (8c,d), when the QP mismatched in stereotypical gender with the pronoun in comparison to conditions (8a,b), when there was a gender match (1025 ms vs. 703 ms).

Discussion

In Experiment 2, the only statistically significant effects we found were main effects of the quantified antecedent's gender, reflecting longer reading times at and after the pronoun region when the stereotypical gender of the QP mismatched in gender with the pronoun. There was also a trend for reading times to be longer when the proper name antecedent mismatched the pronoun in gender compared to when there was a gender match, with the double mismatch condition (8d) eliciting the longest reading times at the spillover region, but this numerical pattern only ever proved marginally significant at best.

Taken together, the results from Experiments 1 and 2 indicate that potential antecedents' proximity to the pronoun, rather than the type of referential dependency involved (variable binding vs. coreference) affected readers' online preferences. Our results suggest that for a QP to be considered during processing, variable binding must be facilitated by additional factors. However, it is conceivable that even though recency influences the likelihood that a QP is initially considered as an antecedent for a pronoun during real-time pronoun resolution, it does not necessarily determine comprehenders' ultimate antecedent choices. In Experiments 1 and 2, we did not directly probe participants' ultimate preferences via post-trial comprehension questions. As such, to examine the extent to which either variable binding or coreference interpretations of ambiguous pronouns are ultimately preferred, we carried out a complementary offline task.

⁵ As in Experiment 1, we also calculated reading times for the two first-pass measures for a *precritical region* containing the text in between the QP antecedent and pronoun in Experiment 2. No reliable differences between conditions were observed for either measure (all $t < 1.38$, all $p > .168$).

Table 3

Reading times for four eye-movement measures at four regions of texts in Experiment 2 (SDs in parentheses).

	First pass reading time	Regression path time	Rereading time	Total viewing time
<i>Pronoun region</i>				
QP Match, Name Match	286 (161)	336 (235)	154 (525)	440 (306)
QP Match, Name Mismatch	301 (175)	408 (387)	195 (417)	496 (456)
QP Mismatch, Name Match	307 (181)	438 (464)	228 (329)	535 (396)
QP Mismatch, Name Mismatch	337 (213)	488 (609)	279 (473)	616 (545)
<i>Spillover region</i>				
QP Match, Name Match	315 (199)	455 (515)	209 (347)	524 (375)
QP Match, Name Mismatch	314 (170)	467 (548)	221 (325)	534 (372)
QP Mismatch, Name Match	309 (191)	491 (733)	237 (316)	546 (368)
QP Mismatch, Name Mismatch	275 (148)	603 (934)	327 (457)	602 (493)
<i>Prefinal region</i>				
QP Match, Name Match	299 (156)	404 (491)	191 (271)	490 (318)
QP Match, Name Mismatch	309 (164)	493 (543)	145 (218)	454 (259)
QP Mismatch, Name Match	306 (184)	440 (424)	184 (280)	490 (340)
QP Mismatch, Name Mismatch	296 (175)	524 (963)	178 (253)	474 (300)
<i>Final region</i>				
QP Match, Name Match	345 (197)	757 (790)	105 (229)	450 (286)
QP Match, Name Mismatch	336 (209)	649 (607)	137 (327)	473 (365)
QP Mismatch, Name Match	356 (253)	945 (1078)	170 (299)	527 (408)
QP Mismatch, Name Mismatch	353 (243)	1104 (1637)	140 (275)	493 (348)

QP = Quantifier Phrase, Name = Proper Name.

Table 4

Summary of statistical analyses for four eye-movement measures at four regions of texts in Experiment 2.

	First pass reading time		Regression path time		Rereading time		Total viewing time	
	Estimate	<i>t</i>	Estimate	<i>t</i>	Estimate	<i>t</i>	Estimate	<i>t</i>
<i>Pronoun region</i>								
QP	25 (14)	1.82*	87 (63)	1.40	67 (37)	1.80*	91 (34)	2.70**
Name	21 (13)	1.68*	69 (51)	1.30	33 (35)	0.95	54 (35)	1.54
QP * Name	12 (28)	0.44	7 (111)	0.06	4 (62)	0.07	4 (63)	0.07
<i>Spillover region</i>								
QP	20 (14)	1.36	89 (60)	1.47	67 (26)	2.57**	45 (27)	1.71*
Name	18 (14)	1.31	61 (57)	1.08	50 (28)	1.77*	30 (27)	1.10
QP * Name	34 (26)	1.29	116 (123)	0.94	84 (70)	1.21	51 (69)	0.74
<i>Prefinal region</i>								
QP	2 (13)	0.17	28 (61)	0.46	22 (18)	1.21	22 (20)	1.12
Name	2 (14)	0.11	75 (57)	1.32	24 (24)	1.00	20 (30)	0.64
QP * Name	21 (33)	0.64	19 (128)	0.15	24 (38)	0.65	2 (42)	0.05
<i>Final region</i>								
QP	12 (25)	0.47	301 (106)	2.84**	34 (21)	1.62	45 (28)	1.62
Name	7 (17)	0.40	30 (110)	0.27	0 (21)	0.02	9 (30)	0.30
QP * Name	7 (38)	0.19	274 (239)	1.15	57 (51)	1.11	52 (55)	0.95

QP = Quantifier Phrase, Name = Proper Name. Estimate = Model Estimate (SE in brackets).

* $p < .10$.** $p < .05$.*** $p < .001$.**Experiment 3**

Experiment 3 was an offline questionnaire in which participants had to choose an antecedent for an ambiguous pronoun. Two conditions were included as in (9a,b), using the experimental materials (excluding the third wrap-up sentence) from the 'double match' conditions (7a) and (8a) in Experiments 1 and 2 respectively.

(9a) *QP – Name*

The squadron paraded through town. Every soldier who knew that James was watching was

convinced that he should wave as the parade passed.

(9b) *Name – QP*

The squadron paraded through town. It looked to James that every soldier was completely convinced that he should wave as the parade passed.

In both (9a) and (9b) the ambiguous pronoun *he* can refer to either *every soldier* or *James*. In both sentences the QP

antecedent *every soldier* is in a structural position such that it c-commands the ambiguous pronoun. In (9a) the QP antecedent is linearly more distant to the pronoun than the proper name antecedent, as was also the case in Experiment 1, whereas in (9b) the linear ordering of antecedents is reversed, as in Experiment 2.

The POB approach predicts that participants should show a preference for the QP antecedent in both (9a,b). A general preference for coreference antecedents, on the other hand, would be reflected in a larger number of proper name choices in both conditions. If the linear order of antecedents affects the extent to which the QP is considered, then we should find the same pattern as we did in our online results, that is, participants should prefer to interpret the pronoun as referring to the proper name in (9a) and the QP in (9b).

Method

Participants

Twenty-four native English speakers (four males, mean age 24) voluntarily took part in Experiment 3, none of whom took part in either Experiments 1 or 2.

Materials

The 24 items from the ‘double match’ conditions in Experiments 1 and 2 were used as in (9). In addition to the 24 experimental items, 48 filler items were also constructed that contained two potential antecedents for a variety of different types of pronouns, some of which were ambiguous, and some of which were unambiguous.

Procedure

The experiment was administered electronically via email as a scalar antecedent choice questionnaire. The pronouns in each text appeared in bold, underlined font and participants were instructed to decide who out of the two potential antecedents they thought each pronoun most likely referred to on a 5-point scale. For (9), a score of 1 or 2 would indicate that the pronoun ‘Very likely refers to *James*’ or ‘Fairly likely refers to *James*’ respectively, while 3 would mean the pronoun was ‘Equally likely to refer to either’, and a score of 4 or 5 would indicate the pronoun ‘Fairly likely refers to *every soldier*’ or ‘Very likely refers to *every soldier*’ respectively.

The experimental and filler items were pseudo-randomised such that no two experimental items occurred next to each other, and were distributed across two lists such that participants only rated one version of each sentence. The questionnaire took approximately 15 min to complete.

Results

Results were coded such that higher scores represented a preference for the QP antecedent. Overall mean scores for the QP – Name condition were 2.64 (*SD* 1.46) and for the Name – QP condition 3.56 (*SD* 1.44). *T*-tests revealed that this difference was significant ($t(23) = 5.61$, $p < .001$, $t(23) = 4.69$, $p < .001$). To examine whether these scores differed from the mid-point of the scale (indicating no clear preference for either antecedent), one sample *t*-tests

compared the mean score of each condition to the value 3. The QP – Name condition received overall ratings significantly lower than 3 ($t(23) = 3.51$, $p < .001$, $t(23) = 2.21$, $p = .004$), while the Name – QP condition was rated significantly higher ($t(23) = 4.65$, $p < .001$, $t(23) = 3.18$, $p < .001$).

These results indicate that participants preferred to link the pronoun to the linearly closest antecedent. In the QP – Name condition, the score of 2.64 suggests a preference for the proper name antecedent, while in the Name – QP condition, the score of 3.56 suggests a preference in the opposite direction, favouring the QP antecedent. Together, the results from Experiment 3 indicate that c-commanding QPs are only preferred if they are linearly the closest antecedent to the pronoun. The results from the offline task will be discussed below, together with the results from our two online experiments.

General discussion

The primary aim of our study was to examine how two alternative pronoun interpretation mechanisms, variable binding and discourse-based coreference assignment, are applied and interact over time. One hypothesis we sought to examine was that variable binding relations should always be computed before coreference assignment, as predicted by the POB economy hierarchy in (4). Our results failed to support this hypothesis however, and instead showed that antecedent recency affected the extent to which a c-commanding, variable binding QP was considered as a potential antecedent for a pronoun.

In two reading experiments and in an offline judgment task, we observed preferences for embedded subject pronouns to refer to the linearly closest of two potential antecedents. In each experiment, one antecedent was a QP that was in a structural position such that it c-commanded the pronoun, while a second antecedent was a proper name. In Experiment 1, when the proper name antecedent did not c-command the pronoun but was linearly closer to it than the QP, we found longer reading times when the proper name antecedent mismatched in gender with the pronoun at and after the pronoun region. We found no reliable effects of the gender of the QP in any reading time measure. In Experiment 2, when the QP antecedent was linearly closer to the pronoun than the proper name, the only reliable effects we found were of the QP’s gender, with longer reading times when the QP mismatched in stereotypical gender with the pronoun. These different effects are exemplified by the total viewing times at the pronoun region in both experiments, as shown in Fig. 1. In Fig. 1(a), we can see that reading times were affected by the gender of the proper name antecedent only. In Fig. 1(b), reading times are longer when the QP mismatched in gender with the pronoun.

Taken together, these results suggest that readers in our experiments attempted to link pronouns to the linearly closest antecedent. In Experiment 1, this led to initially linking the pronoun to the proper name, while in Experiment 2 this led to a preference for the QP antecedent. Experiment 3 indicated that for untimed judgements, the QP was preferred only when it was the linearly closest

antecedent. Below, we discuss the implications of these results for different approaches to pronoun resolution in turn.

Variable binding, coreference, and antecedent recency

The possibility that syntactically mediated dependencies might be easier to compute than those involving discourse-level representations has been articulated in one way or another by a number of different researchers (e.g. Avrutin, 1994, 1999; Burkhardt, 2005; Grodzinsky & Reinhart, 1993). One explicitly articulated framework that makes this prediction is Reuland's (2001, 2011) POB approach to pronoun resolution. The POB model incorporates an economy principle which predicts that variable binding relations should be computed before coreference assignment. While originally formulated as an explanation for interpretive preferences in sentences such as (3) containing VP ellipsis, Koornneef (2008) interpreted the results of an eye-movement experiment with ambiguous pronouns as evidence that the economy principle also guides online pronoun resolution. Note, however, that his experimental design did not allow for particularly fine-grained time-course information, while the gender congruency manipulation we used allowed us to precisely assess the time-course in which each antecedent of an ambiguous pronoun was considered. Charting the time-course of pronoun resolution revealed that contrary to what the economy principle predicts, variable binding is not necessarily initially preferred.

If variable binding relations were always computed before coreference assignment, we should have observed a preference for linking the pronoun to the QP antecedent, as indexed by longer reading times following stereotypical gender violations, in both Experiments 1 and 2. While we observed such an effect in Experiment 2, when the QP was linearly closest to the pronoun, in Experiment 1 we found longer reading times when a coreference antecedent mismatched in gender with the pronoun, at a point in time when we found no evidence of the variable binding antecedent being considered. Our finding that QP antecedents are not always initially preferred is incompatible with the hypothesis that variable binding should always be computed before coreference assignment. Indeed, we found no evidence to suggest that there was always an initial preference for QP antecedents that was later abandoned. Instead, the results of Experiments 1 and 2 indicate that recency affected the extent to which the QP was considered during the initial antecedent search, while the results of Experiment 3 indicate that recency also affected whether a QP was preferred in an untimed judgment task testing comprehenders' preferred final interpretations.

Binding and processing economy

Our finding that syntactically mediated variable binding is not in fact faster than discourse-based coreference assignment might seem surprising in light of earlier findings regarding the online resolution of reflexives. Reflexive binding relations, which are also contingent on c-com-

mand, have been found to be established extremely quickly during processing, even in the presence of a non-c-commanding, intervening competitor antecedent (Sturt, 2003; Xiang et al., 2009).⁶ Within the POB framework, sensitivity to structural relations such as c-command would be expected here under the assumption that reflexive binding is effectively a by-product of independently required syntactic operations during structure-building, in conjunction with argument structure constraints and other lexical properties of predicates (Reuland, 2011).

While the syntactic constraint known as Condition A of the binding theory (Chomsky, 1981) rules out all but one antecedent for a reflexive, non-reflexive pronouns are always technically ambiguous. Note that in our eye-movement experiments, both the proper name and QP antecedents were grammatically licit antecedents of the pronoun. In such cases, when the grammar does not rule out any potential antecedents from the candidate set, pronoun resolution is guided by a number of (potentially interacting) factors, of which c-command information is only one. In our study, the likelihood of a variable binding antecedent being considered during early processing stages depended on its relative proximity to the pronoun. In this way, syntactic cues may 'gate' whether or not other non-structural factors influence retrieval of a variable binding antecedent (see Van Dyke & McElree, 2011, for further discussion of the notion of 'syntactic' gating of memory retrieval during language processing). In short, while reflexive binding relations may indeed be the least computationally costly type of referential dependency, as stated in (4) above, we propose that no intrinsic cost hierarchy or temporal ordering should be imposed on semantic variable binding vs. discourse-based coreference assignment. Recall that Koornneef's (2008) observation that for pronouns in single-antecedent contexts, coreference antecedents led to shorter reading times shortly after the pronoun than variable binding antecedents, is also difficult to reconcile with the POB hypothesis that coreference should always be computationally more costly than variable binding (see also Burkhardt, 2005; Carminati et al., 2002).

The present findings would seem to be consistent with theoretical proposals according to which binding and coreference relations are established at the same representational level, such as the uni-modular approach outlined by Heim (2007). However, in the absence of any clear time-course predictions, it is unclear how alternative semantic approaches to pronoun interpretation can account for our observation that different possible antecedents were preferred depending on their relative position or ordering. It appears that for variable binding to win out over coreference, other factors must come into play that facilitate the computation of variable binding relations such as antecedent recency. For discourse prominence theory (Gordon & Hendrick, 1998) to account for these results, the role of prominence in terms of the height of an ante-

⁶ Although note that the question of whether reflexive binding is entirely immune from interference from structurally illicit antecedents has been the subject of debate (Badecker & Straub, 2002; Cummings & Felser, 2013; Dillon, Mishler, Sloggett, & Phillips, 2013).

cedent in the syntactic tree needs to also interact with antecedent recency.

Memory retrieval during pronoun resolution

In the introduction, we discussed two types of memory search. Serial models in which memory representations are sequentially searched, and content-addressable architectures in which items in memory are accessed directly via a set of retrieval cues. We noted that the POB approach could be implemented as a specific type of serial search in which only antecedents that c-command a subject pronoun are initially retrieved. Alternatively, if prominence is taken to correspond to syntactic salience, similar predictions to the POB approach could potentially be implemented in a content-addressable architecture in which syntactically prominent antecedents have particularly distinct representations in memory. Note that this second approach would largely abandon the main tenet of the POB approach, that c-command determines initial retrieval, and instead reduces c-command to correlated factors related to antecedent prominence. However the predictions of the POB approach are implemented, we found no support for the hypothesis that variable binders *must* be initially retrieved as a virtue of their position in the syntactic structure of a sentence, and as such our results rule out this specific type of search in pronoun resolution.

We initially considered that a number of potential cues could guide antecedent retrieval. One initial hypothesis was that grammatical number and gender agreement cues might guide the search, in which case we predicted that only ‘double mismatch’ conditions (7d/8d) should cause difficulty, as in all others these search criteria can be satisfied by linking the pronoun to either the QP or the proper name. However, this pattern of results was not observed in either experiment. Additionally, our finding of gender mismatch effects across both reading experiments, importantly in experimental conditions when gender-matching antecedents were available, indicates that comprehenders’ may have initially attempted retrieval of a particular antecedent irrespective of gender agreement. This suggests that gender may not be a cue to retrieval, or at the least is not very highly ranked, in itself (see also Dillon et al., 2013), and is also compatible with recent claims that not all cues to memory retrieval during language processing are equally weighted (Van Dyke & McElree, 2011).

Another potential retrieval cue is semantic number, which we argued should favour the proper name antecedent as it provides a more obvious match to the singular number semantics of a singular pronoun than a QP. While the results of Experiment 1 were compatible with this hypothesis, across all experiments we observed a preference for the linearly closest antecedent. A key property of short term memory is that of forgetting, and we believe that the antecedent recency preference we observed could potentially be explained as an example of this phenomenon. However, how to account for forgetting has been widely debated in the memory literature (see Jonides et al., 2008; Van Dyke & Johns, 2012, for review). Below we discuss whether two different theories of forgetting

could account for our observed effects of antecedent recency.

Decay theories of forgetting posit that forgetting results from decreasing levels of activation of items in memory over time. In Lewis et al.’s (2006) model of memory retrieval during sentence comprehension, the activation level of a given item decays over time, which in turn reduces its chances of being subsequently retrieved, all other things being equal. It could be that our observed preference for a variable binding antecedent only when it was linearly the closest antecedent to the pronoun might have resulted from a decay in its level of activation over time when it was more distant, as in Experiment 1. We have reasons to doubt that this be the case, however. Note that in Experiment 1, although the QP was linearly further away from the pronoun than the proper name, it was actually the most recently activated antecedent. For example, for the sentences tested in Experiment 1 (e.g. ‘Every soldier who knew that James was watching was convinced that he...’), although the QP was linearly more distant to the pronoun in the surface structure, it will be retrieved as the subject of the main clause predicate *was convinced*. It is possible that this intermediate retrieval, after the proper name antecedent but before the pronoun, may have increased its activation level as a potential item for subsequent retrieval. However, despite this we still observed a preference to link the pronoun to the proper name antecedent at the pronoun itself. We contend that this would be unexpected if the observed effects were entirely driven by decay over time.

Other accounts claim that forgetting is primarily a result of interference from other items in memory. In feature-overwriting models, forgetting is explained in terms of interference from items in memory that contain similar content (e.g. Nairne, 1990; Oberauer & Kliegl, 2006).⁷ Feature-overwriting occurs when memories overlap in content. This leads to less distinct representations in memory for the items in question, which in turn leads to a decrease in the probability of them being successfully retrieved. If forgetting over time is explained in terms of more recently encoded items having more distinct representations in memory than those previously encoded, it is possible that the most recently encoded antecedent in our experiments had a more distinct representation in memory than the linearly more distant one. In this way, our observation that the QP was preferred only when it was the linearly closest antecedent, can be explained as resulting from the fact that the QP will have a more distinct representation in memory when it was the most recently encoded item, thus increasing the probability of it being retrieved, other things being equal.⁸ The hypothesis that interference as a result of feature-overwriting can explain the observed antecedent recency effect is also in line with recent proposals that the primary determinant of forgetting is interference, rather than decay (e.g.

⁷ We thank an anonymous reviewer for pointing out how feature-overwriting might explain our results.

⁸ Within this kind of scenario, the marginal trend for the more distant proper name antecedent to influence first-pass times at the pronoun seen in Experiment 2 might have been due to proper names having particularly stable discourse representations (e.g. Sanford & Garrod, 1988).

Berman, Jonides, & Lewis, 2009). Note, however, that the results of Experiment 1 indicate that it is the linear position in which an antecedent is initially encoded that seems to drive the recency effect. Potentially related effects have also been observed in filler-gap dependencies. For example, McElree et al. (2003) observed that asymptotic accuracy in a speed-accuracy tradeoff task was influenced by the distance between the ultimate gap site and the linear surface position in which the filler was initially encoded, irrespective of intervening sites of intermediate retrieval.

Although antecedent recency can provide an account of why the c-commanding QP was preferred only when it was linearly the closest antecedent to the pronoun, we do not claim recency to be the only relevant factor in antecedent retrieval. Precisely how variable binding, coreference assignment and antecedent recency interact with other factors that are known to influence anaphora resolution, such as implicit causality (Garvey & Caramazza, 1974), subjecthood (Järvikivi et al., 2005) and parallelism (Chambers & Smyth, 1998) amongst many others, is open to future research. As a reviewer points out, participants in our experiments might also have been influenced by subtle discourse biases, such as the suggestion that in Experiment 2, James appears to have an attitude about every soldier, whereas in Experiment 1, some soldiers might be said to have an attitude about James. Additionally, in the materials used in Experiment 2, the QP was assigned the semantic role of agent while the proper name antecedent was an experiencer. If semantic roles cue antecedent retrieval, and agents are generally preferred, then this could have been an additional factor, on top of recency, that may have favoured the QP antecedent. Future research will be required to fully tease apart under what conditions a variable binding or coreference antecedent is preferentially retrieved during anaphora resolution. The primary aim of the current paper was however to investigate whether a particular type of pronoun-antecedent relation was always initially preferred, and the results of the present study clearly show that, in contrast to the economy principle in (4), variable binding relations are not always computed first.

One interesting empirical question is whether a non c-commanding QP will ever be considered as a potential antecedent for a pronoun. This option should be ruled out if grammatical cues act as ‘hard constraints’ to antecedent retrieval in a structure-sensitive search (Dillon et al., 2013), but not necessarily in a direct access search in which interference can be observed from items in memory that are nevertheless grammatically unavailable (Van Dyke, 2007). Carminati et al. (2002) provided some evidence that variable binding interpretations are possible for quantified phrases that do not c-command a pronoun. However, they argued that interpretations similar to variable binding in the contexts that they studied were possible by a phenomenon known as ‘telescoping’ (Roberts, 1989), in which it is possible to infer a set of referential antecedents for a quantified antecedent that provide a similar interpretation to ‘true’ cases of bound variable anaphora (Carminati et al., 2002: 22; see also Bosch, 1983). Preliminary results from more recent work suggest that c-command does play an important role in

determining whether variable binders are considered as antecedents for an ambiguous pronoun (Kush, 2013). Further investigation is required to examine the precise structural configurations in which QPs are considered, or ignored, as potential antecedents of a pronoun during processing.

Conclusion

In three experiments we investigated whether ambiguous pronouns are preferentially resolved via either variable binding or coreference assignment. We found no overall preference for either route to pronoun resolution, and instead observed that comprehenders only preferred to link a pronoun to a variable binding QP when it was the linearly closest antecedent. These findings were interpreted as being incompatible with theories of pronoun resolution which predict that one particular route to pronoun interpretation should always be favoured initially, and in particular are incompatible with the hypothesis that variable binding relations should always be computed before coreference assignment. Instead, our results are compatible with a memory search architecture in which forgetting over time results in more recently encoded items having more distinct representations. Taken together, the results of the current study indicate that for variable binding antecedents to be considered during the early stages of pronoun resolution when multiple antecedents are grammatically available, the computation of variable binding relations must be facilitated by additional factors such as antecedent recency.

Acknowledgments

We thank the editor and three anonymous reviewers for their suggestions and insightful comments on previous drafts of this paper. The research reported here and preparation of the manuscript was supported by a British Academy Postdoctoral Fellowship (pf100026) to I.C., and an ESRC Postgraduate Studentship to C.P. Part of the research reported here was also funded by an Alexander-von-Humboldt Professorship awarded to Harald Clahsen.

Appendix

The materials for Experiments 1 and 2 are given below. The QP and proper name antecedent manipulations are shown in parentheses. Square brackets denote the antecedent linear order manipulations in Experiments 1 and 2, delimited with a forward slash (/).

1. The squadron paraded through town. [Every soldier who knew that James (Helen) was watching was convinced/It looked to James (Helen) that every soldier was completely convinced] that he (she) should wave as the parade passed. The entire town was extremely proud that day.
2. The new construction site was busy. [Every builder who saw that Steven (Joanna) was at the site was told/It appeared to Steven (Joanna) that every

- builder was emphatically told] that he (she) should complete a full day's work. There was no time for any slacking off!
3. The building was unsafe. [Every firefighter who believed that John (Jane) was inside was informed/It appeared to John (Jane) that every firefighter was very quickly informed] that he (she) must wait for more help to arrive. Luckily nobody was seriously injured that day.
 4. It was another day of training. [Every footballer who saw that Roger (Sarah) was on the pitch was told/It appeared to Roger (Sarah) that every footballer was very specifically told] that he (she) had to practice for 3 h. It was going to be a tough workout for sure.
 5. It was very busy at the airport. [Every pilot who knew that David (Diana) was passing through was relieved/It seemed to David (Diana) that every pilot was certainly very relieved] that he (she) did not get at all delayed. Air traffic control were running things very efficiently.
 6. The boxing night was always popular. [Every boxer who heard that Adam (Emma) was at the event knew/It looked to Adam (Emma) that every boxer almost definitely knew] that he (she) needed to really impress the crowd. It was certainly not going to be very easy.
 7. A farming trade fair was in town. [Every butcher who heard that Bob (Ann) was at the fair was informed/It seemed to Bob (Ann) that every butcher was quite happily informed] that he (she) would be able to make a speech. The day was considered to be a great success by all.
 8. The large passenger ferry left port. [Every sailor who saw that Mark (Mary) was on deck was informed/It appeared to Mark (Mary) that every sailor was dutifully informed] that he (she) should be ready for the trip ahead. There was a chance that the sea was going to be rough.
 9. The garage was full of cars. [Every mechanic who knew that Peter (Susan) was being lazy was surprised/It looked to Peter (Susan) that every mechanic was most happily surprised] that he (she) could have a long lunch break. Some people seem to have all the luck!
 10. The church had a big congregation. [Every priest who heard that Paul (Katy) was newly ordained agreed/It seemed to Paul (Katy) that every priest almost immediately agreed] that he (she) should talk to the local people. It was important that [Everybody knew each other.
 11. The mining village was a lively place. [Every miner who knew that Bill (Lucy) was hard at work was told/It appeared to Bill (Lucy) that every miner was enthusiastically told] that he (she) could get some time off next week. It was something worth looking forward to.
 12. Good training is very important. [Every electrician who heard that Ben (Amy) was on a course thought/It looked to Ben (Amy) that every electrician understandably thought] that he (she) should work extra hours each week. It is essential to get as much work experience as possible.
 13. People can be superstitious. [Every fortune teller who knew that Helen (James) was a believer thought/It seemed to Helen (James) that every fortune teller most definitely thought] that she (he) should make plans about the future. It is always important to try and think ahead.
 14. The delivery was due very soon. [Every florist who knew that Joanna (Steven) was visiting the shop was told/It looked to Joanna (Steven) that every florist was quite assuredly told] that she (he) would like all the new stock. It is important to have a variety of products on sale.
 15. Some jobs are not well paid. [Every babysitter who believed that Jane (John) was underpaid was reassured/It seemed to Jane (John) that every babysitter was delightfully reassured] that she (he) would be given a pay rise soon. That was certainly some good news to hear.
 16. It was another day on the ward. [Every midwife who noticed that Sarah (Roger) was at the hospital thought/It looked to Sarah (Roger) that every midwife almost undoubtedly thought] that she (he) should prepare for a busy day. There is always much to be done at the maternity unit.
 17. Life at work can be hectic. [Every typist who understood that Diana (David) was in a hurry decided/It appeared to Diana (Diana) that every typist very promptly decided] that she (he) should try and work a little faster. It was getting late and Everyone wanted to go home.
 18. The office was very competitive. [Every secretary who heard that Emma (Adam) was visiting was reminded/It seemed to Emma (Adam) that every secretary was continually reminded] that she (he) could get a promotion very soon. It was going to be an interesting month at work.
 19. The country mansion was beautiful. [Every housekeeper who saw that Ann (Bob) appreciated the garden hoped/It appeared to Ann (Bob) that every housekeeper very wholeheartedly hoped] that she (he) would be able to enjoy the day. The weather did look like it was going to be warm and sunny.
 20. The beauty salon was very popular. [Every beautician who saw that Mary (Mark) was passing by was pleased/It looked to Mary (Mark) that every beautician was really rather pleased] that she (he) could have a chat about work. It made the day a little bit more interesting.
 21. It was very busy at the hospital. [Every nurse who noticed that Susan (Peter) was on the ward was relieved/It seemed to Susan (Peter) that every nurse was quite thoroughly relieved] that she (he) could go home an hour early. It had been a very long and tiring day.
 22. It was a competitive business. [Every fashion model who saw that Katy (Paul) was new was reassured/It seemed to Katy (Paul) that every fashion model was quickly reassured] that she (he) would not be given special treatment. There was definitely some tension in the dressing room.

23. It was a tough schedule. [Every cheerleader who noticed that Lucy (Bill) was having trouble said/It appeared to Lucy (Bill) that every cheerleader very emphatically said] that she (he) needed some time off from training. The chance to take a rest was certainly welcome.
24. There was much housework to be done. [Every cleaner who heard that Amy (Ben) was on holiday was informed/It seemed to Amy (Ben) that every cleaner was rather quietly informed] that she (he) would work more hours next week. It was the busiest time of the entire year.

References

- Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. C. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, 22, 527–565.
- Avrutin, S. (1994). *Psycholinguistic investigations in the theory of reference*. Doctoral thesis, MIT.
- Avrutin, S. (1999). *Development of the syntax–discourse interface*. Dordrecht: Kluwer Academic.
- Baayen, H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baayen, H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.
- Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 748–769.
- Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Berman, M., Jonides, J., & Lewis, R. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 35, 317–333.
- Bosch, P. (1983). *Agreement and anaphora: A study of the role of pronouns in syntax and discourse*. London: Academic Press.
- Burkhardt, P. (2005). *The syntax–discourse interface*. Amsterdam & Philadelphia: John Benjamins.
- Burzio, L. (1998). Anaphora and soft constraints. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, & D. Pesetsky (Eds.), *Is the best good enough? Optimality and competition in syntax*. Massachusetts: MIT Press.
- Carminati, M., Frazier, L., & Rayner, K. (2002). Bound variables and c-command. *Journal of Semantics*, 19, 1–34.
- Chambers, C., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language*, 39, 593–608.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Collins, C. (2005). A smuggling approach to raising in English. *Linguistic Inquiry*, 36, 289–298.
- Cooreman, A., & Sanford, A. (1996). *Focus and syntactic subordination in discourse*. Technical report RP-79. Human Communication Research Centre, University of Edinburgh.
- Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, 28, 188–219.
- Dillon, B. (2011). *Structured access in sentence comprehension*. Unpublished PhD thesis, University of Maryland.
- Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modelling evidence. *Journal of Memory and Language*, 69, 85–103.
- Evans, G. (1980). Pronouns. *Linguistic Inquiry*, 11, 337–362.
- Foley, C., Nunez del Prado, Z., Barbier, I., & Lust, B. (2003). Knowledge of variable binding in VP ellipsis: Language acquisition research and theory converge. *Syntax*, 6, 52–83.
- Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Availability vs. accessibility. *Journal of Memory and Language*, 56, 357–383.
- Frazier, L., & Clifton, C. (2000). On bound variable interpretations: The LF-only hypothesis. *Journal of Psycholinguistic Research*, 29, 125–140.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5, 459–464.
- Gernsbacher, M. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, 27, 699–717.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 311–347.
- Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science*, 22, 389–424.
- Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and of coreference. *Linguistic Inquiry*, 24, 69–101.
- Heim, I. (1992). *The semantics of definite and indefinite noun phrases*. University of Massachusetts: Ph.D. dissertation; published 1989. New York: Garland.
- Heim, I. (1998). Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach. In Sauerland, G., & Percus, O. (Eds.), *The interpretive tract. MIT working papers in linguistics* (Vol. 25, pp. 205–246). MIT: Cambridge, MA.
- Heim, I. (2007). Forks in the road to rule I. In M. Abdurrahman, A. Schardl, & M. Walkow (Eds.), *Proceedings of the thirty-eighth annual meeting of the North East Linguistic Society*. University of Ottawa (26–28 October 2007).
- Hendriks, P., & Spender, J. (2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition*, 13, 319–348.
- Järviski, J., van Gompel, R., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, 16, 260–264.
- Jonides, J., Lewis, R., Nee, D., Lusting, C., Berman, M., & Moore, K. (2008). The mind and brain of short term memory. *Annual Review of Psychology*, 59, 193–224.
- Kazanina, N., Lau, E., Liberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backward anaphora. *Journal of Memory and Language*, 56, 384–409.
- Koornneef, A. (2008). *Eye-catching anaphora*. Utrecht: LOT International Dissertation Series.
- Koornneef, A., Wijnen, F., & Reuland, E. (2006). Towards a modular approach to anaphora processing. In R. Artstein, & M. Poesio (Eds.), *Ambiguity in Anaphora workshop proceedings* (pp. 65–72). Malaga: ESSLLI.
- Koornneef, A., Avrutin, S., Wijnen, F., & Reuland, E. (2011). Tracking the preference for bound-variable dependencies in ambiguous ellipses and only-structures. In J. Runner (Ed.), *Experiments at the interfaces* (Syntax and Semantics, 37). Bingley, UK: Emerald Group Publishing Limited.
- Koornneef, A. (2010). Looking at anaphora: The psychology reality of the Primitives of Binding model. In T. Everaert, H. Lentz, Ø. de Mulder, A. Nilsen, & A. Zondervan (Eds.), *The linguistic enterprise: From knowledge of language to knowledge of linguistics* (pp. 141–166). Amsterdam: John Benjamins.
- Kush, D. (2013). *Respecting relations: Memory access and antecedent retrieval in incremental sentence processing*. Doctoral thesis, University of Maryland.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10, 44–54.
- Martin, A., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb–phrase ellipsis. *Journal of Memory and Language*, 58, 879–906.
- McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, 29, 111–123.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67–91.
- Nairne, J. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251–269.
- Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55, 601–626.
- Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell.
- Rayner, K. (1998). Eye-movements in reading and information processing: 20 Years of research. *Psychological Bulletin*, 124, 372–422.
- Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.
- Reuland, E. (2001). Primitives of binding. *Linguistic Inquiry*, 32, 439–492.
- Reuland, E. (2011). *Anaphora and language design*. Cambridge, Massachusetts: The MIT Press.
- Roberts, C. (1989). Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy*, 12, 683–721.

- Sanford & Garrod (1988). Proper names as controllers of discourse focus. *Language and Speech*, 31, 43–56.
- Streb, J., Henninghausen, E., & Rösler, F. (2004). Different anaphoric expressions are investigated by event-related brain potentials. *Journal of Psycholinguistic Research*, 33, 175–201.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.
- Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 33, 407–430.
- Van Dyke, J., & Johns, C. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass*, 6, 193–211.
- Van Dyke, J., & McElree, B. (2006). Retrieval interference in sentence processing. *Journal of Memory and Language*, 55, 157–166.
- Van Dyke, J., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65, 247–263.
- Van Gompel, R., & Liversedge, S. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 128–139.
- Vasić, N., Avrutin, S., & Ruigendijk, E. (2006). Interpretation of pronouns in VP-ellipsis constructions in Dutch Broca's and Wernicke's aphasia. *Brain and Language*, 96, 191–206.
- Wagers, M., Lau, E., & Phillips, C. (2009). Attraction in comprehension: Representation and processes. *Journal of Memory and Language*, 61, 206–237.
- Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain & Language*, 108, 40–55.