



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

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

Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence

Brian Dillon^{a,*}, Alan Mishler^b, Shayne Sloggett^a, Colin Phillips^c

^a Department of Linguistics, University of Massachusetts, Amherst, USA

^b Center for Advanced Study of Language, MD, USA

^c Department of Linguistics, Program in Neuroscience and Cognitive Science, University of Maryland, USA

ARTICLE INFO

Article history:

Received 16 December 2011
revision received 6 April 2013
Available online 14 May 2013

Keywords:

Agreement
Reflexives
Sentence processing
Memory retrieval
ACT-R

ABSTRACT

We investigated the relationship between linguistic representation and memory access by comparing the processing of two linguistic dependencies that require comprehenders to check that the subject of the current clause has the correct morphological features: subject–verb agreement and reflexive anaphors in English. In two eye-tracking experiments we examined the impact of structurally illicit noun phrases on the computation of reflexive and subject–verb agreement. Experiment 1 directly compared the two dependencies within participants. Results show a clear difference in the intrusion profile associated with each dependency: agreement resolution displays clear intrusion effects in comprehension (as found by Pearlmuter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2009), but reflexives show no such intrusion effect from illicit antecedents (Sturt, 2003; Xiang, Dillon, & Phillips, 2009). Experiment 2 replicated the lack of intrusion for reflexives, confirming the reliability of the pattern and examining a wider range of feature combinations. In addition, we present modeling evidence that suggests that the reflexive results are best captured by a memory retrieval mechanism that uses primarily syntactic information to guide retrievals for the anaphor's antecedent, in contrast to the mixed morphological and syntactic cues used to resolve subject–verb agreement dependencies. Despite the fact that agreement and reflexive dependencies are subject to a similar morphological agreement constraint, in online processing comprehenders appear to implement this constraint in distinct ways for the two dependencies.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Language comprehension clearly relies on working memory systems to construct the relationships between words and phrases that allow comprehenders to understand sentences and discourses. Long-distance dependencies—linguistic relationships that can span a large amount of intervening material—have long been a focus

of research into the nature of these memory systems (e.g., Chomsky & Miller, 1963, chap. 11; Gibson, 1998; Lewis & Vasishth, 2005). In order to gain insight into the memory systems that support sentence comprehension, researchers have generally focused on linguistic configurations that are difficult or error-prone. In recent years, research in this tradition has provided an increasing amount of experimental support for a content-addressable memory system as a crucial component of the parser (McElree, 2000; McElree, Foraker, & Dyer, 2003). The defining characteristic of a content-addressable memory architecture is that it allows direct access to memory representations whose content matches a set of retrieval

* Corresponding author. Address: Department of Linguistics, 226 South College, University of Massachusetts, 150 Hicks Way, Amherst, MA 01003, USA.

E-mail address: brian@linguist.umass.edu (B. Dillon).

cues. This retrieval mechanism stands in contrast to retrieval mechanisms that engage a serial search through memory in that it predicts the time course of memory access should be constant across different dependency lengths. McElree (2000) and McElree et al. (2003) showed that memory retrieval for a number of linguistic dependencies had this property (see also Foraker & McElree, 2007; Martin & McElree, 2008, 2009). In such an architecture, however, memory access may be impeded if multiple representations match the retrieval cues, wholly or in part. Thus an additional source of experimental support for this memory access mechanism comes from the predictions that it makes about interference effects in memory retrieval (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke, 2007; Vasishth, Brüssow, Lewis, & Drenhaus, 2008).

More generally this line of work has been taken to support the claim that a single set of general principles governs memory retrieval in sentence processing tasks (Lewis et al., 2006). In particular, this view suggests that a single, direct-access retrieval mechanism is used for processing linguistic relations as diverse as subject–verb agreement, filler–gap processing, licensing of negative-polarity items (NPIs), verb–phrase ellipsis, and referential processing (Foraker & McElree, 2007; Lewis & Vasishth, 2005; Lewis et al., 2006; Martin & McElree, 2008, 2009; McElree et al., 2003; Vasishth et al., 2008; Wagers, Lau, & Phillips, 2009). In order to understand how such a mechanism can accomplish the task of sentence comprehension, it is necessary to specify the retrieval cues used to access linguistic memory. Thus a major question for cue-based parsing models concerns the nature of the cues used and how they are combined at retrieval (see Van Dyke & McElree, 2011). This question is especially relevant when one considers the range of linguistic dependencies listed above, each of which is associated with its own grammatical and interpretive constraints that may restrict memory retrieval. In light of this variation, an important question concerns the relationship between the grammatical constraints associated with a given dependency and the cues that are used to guide memory retrievals associated with that dependency.

One possible answer to this question is that there is a simple mapping between overt linguistic features and retrieval cues: all features associated with a dependency are deployed as retrieval cues for the purposes of parsing-related memory retrievals. For example, in English the verb and its subject must agree in person and number features, and so these morphological features would provide effective retrieval cues if the subject is to be retrieved for integration with the verb (Vasishth et al., 2008; Wagers, 2008; Wagers et al., 2009). If this simply reflects the correlation between the overt morphological features of the verb and the features of the target subject, then any linguistic dependencies with superficially identical agreement constraints should show similar retrieval profiles. On this account, both English subject–verb agreement and reflexive–antecedent dependencies should use morphological features as retrieval cues, as they both require agreement between elements in the dependency: for example, *the boys were angry* and *the boys hit themselves*

are grammatical, but **the boys was angry* and **the boys hit himself* are ungrammatical. A similar view is endorsed by models in which cue validity—the overall usefulness of a cue—determines its use in parsing, perhaps in proportion to its overall frequency or its reliability in the language in question (Bates & MacWhinney, 1989).

However, an alternative view is that retrieval cues for parsing-related retrievals reflect the grammatical function of the dependencies being processed. This view suggests that overt correlations between agreement features are not sufficient for morphological features to be used as retrieval cues. For example, although subject–verb agreement and reflexive–antecedent agreement both require morphological feature concord, the source of this agreement constraint is arguably different for the two dependencies. For reflexive–antecedent relations, agreement reflects a referential dependency between two noun phrases (Baker, 2008; Kratzer, 2009), whereas subject–verb agreement is a formal morphosyntactic mechanism for indexing participants in an event (Corbett, 2006). These differences in the source of the agreement relation appear to correlate with differences in grammatical behavior (Baker, 2008). On these models, one might expect morphological features not to be used as cues for retrievals that construct the reflexive binding dependency, as these features are not directly implicated in the grammatical binding relation. On the other hand, agreement dependencies are by definition grammatical dependencies between the morphosyntactic feature content on two elements in a parse, and so on this view it is natural to implement these constraints as retrieval cues in a cue-based parsing architecture. It is worth noting that models that posit different retrieval mechanisms for various linguistic dependencies (Dillon, 2011; Phillips, Wagers, & Lau, 2011) predict that distinct grammatical dependencies should show distinct patterns of sensitivity to different linguistic features at the point of retrieval.

In the present study, we ask whether the use of morphological agreement features in retrieval reflects a correlation between the overt features of two elements in a dependency, or if the grammatical function of a dependency determines the use of cues for retrieval. To this end, we contrast the processing of reflexive–antecedent and subject–verb agreement dependencies in English. The important feature of these dependencies is that they both require retrieval of the local subject, and they both require the local subject to carry the same agreement features as the dependent element (the agreeing verb or reflexive). If there is a uniform mapping from surface linguistic features to retrieval cues, then these two dependencies should show identical intrusion profiles. Chen, Jaeger, and Vasishth (2011) and Patil, Vasishth, and Lewis (2011) advocate this view, arguing from studies of English and Chinese reflexives that these anaphors use agreement feature cues to retrieval in just the same way that subject–verb agreement does. On the other hand, if the use of agreement features in retrieval is a determined by the grammatical function of a dependency, then we do not expect subject–verb agreement and reflexive–antecedent dependencies to recruit agreement features as retrieval cues in the same fashion.

Intrusion effects

In order to evaluate whether particular linguistic features are used as retrieval cues, a natural place to look is for interference effects at the point of the hypothesized memory retrieval. In a content-addressable architecture, if a memory representation contains features that overlap with the retrieval cues, then it resonates with those features and may be made available for further processing. When multiple memories match the retrieval cues, then retrieval interference occurs and processing difficulty results. This sort of similarity-based interference has been observed for a range of linguistic dependencies, such as subject–verb integration (Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, Johnson, & Lee, 2006; Gordon, Hendrick, & Levine, 2002; Van Dyke, 2007; Van Dyke & Lewis, 2003) and filler–gap relations (Van Dyke & McElree, 2006).

However, in a number of cases memory interference results in processing facilitation (sometimes called *intrusion*). There are two linguistic dependencies where this has been observed to date: subject–verb agreement and negative polarity item (NPI) processing (Clifton, Frazier, & Deevy, 1999; Drenhaus, Frisch, & Saddy, 2005; Pearlmutter, Garnsey, & Bock, 1999; Vasishth et al., 2008; Wagers et al., 2009; Xiang, Dillon, & Phillips, 2009). Across these studies, facilitation is seen in ungrammatical sentences that have syntactically illicit feature-matched elements, relative to ungrammatical controls that lack a feature-matched element. This facilitation is sometimes understood as reflecting a retrieval error that makes an ungrammatical sentence appear well-formed, and so facilitatory interference effects have been referred to as illusions of grammaticality (Phillips et al., 2011). As noted by Vasishth et al. (2008) and Wagers et al. (2009), intrusion effects in comprehension are an important prediction of content-addressable memory architectures. In particular, these facilitatory effects provide relatively direct evidence that comprehenders retrieve and consider structurally illicit elements due to cue overlap. This stands in contrast to similarity-based interference, which could in principle arise as a function of distinctiveness in memory at the point of retrieval, a point we return to in the discussion.

Intrusion effects for subject–verb agreement are the comprehension analog of the well-known agreement attraction effect in production, which has been noted by a wide range of researchers (e.g., Bock & Miller, 1991; Solomon & Pearlmutter, 2004; Vigliocco, Butterworth, & Garrett, 1996; Vigliocco & Nicol, 1998; for grammatical treatments, see den Dikken, 2001; Kayne, 1989; Kimball & Aissen, 1971). Agreement attraction occurs when the morphological features of a noun other than the local subject appear to control verbal agreement. For example, Bock and Miller (1991) presented participants with sentence fragments such as *the key to the cabinet(s)*. They observed that when the distractor noun *cabinet* was plural, there was a marked increase in the probability that participants would produce a plural verb form (i.e. *were*), relative to conditions where the distractor was singular. The agreement attraction effect has been shown to obtain in a range of environments and across a range of different types of

distractor noun (see Eberhard, Cutting, & Bock, 2005 for a review).

The impact of syntactically illicit noun phrases on the computation of agreement is also observed in comprehension measures (Clifton et al., 1999; Pearlmutter et al., 1999; Wagers et al., 2009). Across a number of studies, a clear generalization about the comprehension profile of agreement intrusion has emerged. For instance, Wagers et al. (2009) employed self-paced reading to investigate the processing of grammatical and ungrammatical agreement dependencies. They examined sentences of the form in (1):

-
- (1) a. The key to the cabinet(s) unsurprisingly was rusty from years of disuse.
 b. *The key to the cabinet(s) unsurprisingly were rusty from years of disuse.
-

They consistently found no difference in the grammatical conditions in (1a). However, in the ungrammatical conditions in (1b), a plural distractor NP *cabinets* led to faster reading times at the critical verb, relative to an ungrammatical condition with a singular distractor. In other words, having a syntactically illicit, but feature-matched NP led to processing facilitation in these sentences. This illusion of grammaticality is robustly observed under different methodologies and syntactic configurations. Pearlmutter et al. (1999) reliably found this effect across a number of experiments with both self-paced reading and eye-tracking measures.

Wagers and colleagues argued that cue-based retrieval interference was responsible for the intrusion effects in their study. In particular they proposed that when the distractor noun had the morphological features needed to license the verb, it was occasionally erroneously retrieved due to the overlap with the morphological retrieval cues. Because of the incorrect retrieval of the distractor NP, comprehenders sometimes perceived the incorrect agreement to be grammatical, leading to faster processing and higher rates of acceptance for plural distractor noun conditions in (1b), compared to singular distractor noun conditions. The most compelling piece of evidence for this, according to these authors, is the grammatical asymmetry prediction of the retrieval account. Wagers and colleagues argued that because the grammatical conditions contain a perfectly feature-matched target NP (the subject head noun), a distractor that only partially matches the retrieval cues is less likely to intrude. Their data supported this prediction, providing support for a cue-based retrieval explanation of the attraction effect in comprehension.

The limits of intrusion in comprehension

Intrusion effects in agreement comprehension provide evidence for the use of agreement features as retrieval cues. Interestingly, studies that have examined reflexive comprehension have failed to find clear evidence for any such intrusion effect in this dependency. A number of authors have argued that this pattern suggests that

syntactic structure provides a hard constraint on the nouns that are initially considered for participation in a reflexive dependency; in other words, these results suggest that only syntactic cues are used to retrieve the target antecedent (Clackson, Felser, & Clahsen, 2011; Clifton et al., 1999; Nicol & Swinney, 1989; Sturt, 2003; Xiang et al., 2009; but cf. Badecker & Straub, 2002).

For example, Sturt (2003) used eye-tracking to determine whether a morphological feature match with structurally illicit antecedents influenced the construction of the antecedent–reflexive dependency. He examined two types of mini-discourse, one of which is shown in (2). In these materials the gender-biased nouns (e.g., *the surgeon*) lead comprehenders to commit to the gender of the subject (Osterhout, Bersick, & McLaughlin, 1997), causing a temporary anomaly (marked here with #) when the local subject's stereotypical gender mismatches the morphological features of the reflexive.

-
- (2) Jonathan was pretty worried at the City Hospital.
- a. The surgeon who treated Jonathan had pricked himself with a needle.
 - b. #The surgeon who treated Jonathan had pricked herself with a needle.
Jennifer was pretty worried at the City hospital.
 - c. The surgeon who treated Jennifer had pricked himself with a needle.
 - d. #The surgeon who treated Jennifer had pricked herself with a needle.
-

Across two experiments, Sturt failed to find evidence for an intrusion effect. In almost all measures and configurations, comprehenders were sensitive only to the feature match between the reflexive and the appropriate local subject. The only exception to this generalization was an effect of the illicit antecedent observed in re-reading times in Sturt's Experiment 1. In this measure, a feature-matched illicit antecedent was associated with faster reading times than an unmatched antecedent at the reflexive in grammatical environments. However, this effect did not replicate in his Experiment 2, and the direction of the numerical trend was reversed. The lack of intrusion for reflexives in similar stimuli was replicated by Xiang et al. (2009) using event-related potentials (ERPs).

Experiment 1

The apparent contrast between agreement and reflexives is valuable for evaluating the relationship between linguistic constraints and retrieval cues. However, it is difficult to compare the intrusion profiles of agreement and reflexive retrievals based on the studies reported above, as these studies employed different experimental designs and methodologies (see, e.g., Clifton et al., 1999). In order to investigate any potential contrast in intrusion profiles between reflexives and agreement, Experiment 1 made a within-subjects comparison of subject–verb agree-

ment and reflexive licensing in English using eye-tracking. For both dependencies the question is whether syntactically illicit, feature-matched noun phrases impact the formation of the dependency at the point of retrieval of the local subject. This retrieval is triggered at the inflected verb in subject–verb agreement dependencies, and at the reflexive anaphor for reflexive–antecedent dependencies. Of particular interest is whether or not the intrusion profile commonly observed for agreement also obtains for reflexives. If so, then we should observe an illusion of grammaticality effect for reflexives: ungrammatical sentences with feature-matched distractors should be easier to process than ungrammatical sentences without a feature-matched distractor, as reflected in faster reading times or fewer regressions.

To make this comparison, we controlled two linguistic factors that have been confounded with dependency across previous online studies of reflexives and verbal agreement. The first such factor is the syntactic environment of the distractor. In all conditions the distractor was embedded inside a subject relative clause (RC), rather than the PP modifier that has been used in most previous agreement intrusion studies. The second factor that was matched was the morphological feature manipulation on the distractor noun. Studies on reflexive processing have investigated the impact of gender features on the distractor, while studies of agreement have focused on number. Because English verbal morphology only marks contrasts in number we tested the impact of matching number features on the formation of reflexive dependencies. The subject noun phrase was held constant across all conditions, differing only in the morphological features of the distractor NP.

Participants

Forty members of the University of Maryland community participated in Experiment 1 (24 females, mean age 21.9). Participants gave informed consent, and were either paid \$10 for their participation or received course credit. The experimental session, including set-up and calibration, lasted approximately 1 h.

Stimuli

Two experimental factors were manipulated in a parallel fashion within each of the two dependencies: grammaticality and intrusion. Grammaticality was manipulated by varying the number feature of the agreeing verb or reflexive. The agreeing element was singular in the grammatical conditions, and plural in the ungrammatical conditions. Likewise, intrusion was manipulated by varying the number of the distractor NP. Based on previous findings (Bock & Miller, 1991; Eberhard et al., 2005; Pearlmuter et al., 1999; Wagers et al., 2009), singular distractor nouns were predicted to cause no intrusion, whereas plural embedded nouns were potential sources of intrusion. The head noun was always singular. These factors were fully crossed within each dependency.

Forty-eight item sets of the form shown in Tables 1 and 2 were constructed. For agreement and reflexive dependencies alike the subject noun phrase was held constant. In

Table 1

Summary of agreement conditions in Experiment 1. Critical and spillover regions included in the analysis are underlined.

Agreement conditions for Experiment 1
<i>Grammatical, no intrusion</i>
1. <u>The new executive/</u> who oversaw/ <u>the middle manager/</u> apparently/ <u>was dishonest/</u> <u>about the company's profits</u>
<i>Grammatical, intrusion</i>
2. <u>The new executive/</u> who oversaw/ <u>the middle managers/</u> apparently/ <u>was dishonest/</u> <u>about the company's profits</u>
<i>Ungrammatical, no intrusion</i>
3. <u>The new executive/</u> who oversaw/ <u>the middle manager/</u> apparently/ <u>were dishonest/</u> <u>about the company's profits</u>
<i>Ungrammatical, intrusion</i>
4. <u>The new executive/</u> who oversaw/ <u>the middle managers/</u> apparently/ <u>were dishonest/</u> <u>about the company's profits</u>

Table 2

Summary of reflexive conditions in Experiment 1. Critical and spillover regions included in the analysis are underlined.

Reflexive conditions for Experiment 1
<i>Grammatical, no intrusion</i>
1. <u>The new executive/</u> who oversaw/ <u>the middle manager/</u> apparently doubted/ <u>himself on/</u> <u>most major decisions</u>
<i>Grammatical, intrusion</i>
2. <u>The new executive/</u> who oversaw/ <u>the middle managers/</u> apparently doubted/ <u>himself on/</u> <u>most major decisions</u>
<i>Ungrammatical, no intrusion</i>
3. <u>The new executive/</u> who oversaw/ <u>the middle manager/</u> apparently doubted/ <u>themselves on/</u> <u>most major decisions</u>
<i>Ungrammatical, intrusion</i>
4. <u>The new executive/</u> who oversaw/ <u>the middle managers/</u> apparently doubted/ <u>themselves on/</u> <u>most major decisions</u>

all cases the subject head noun (NP1) was modified by a subject relative clause that contained the distractor noun (NP2). In order to ensure that the overt gender marking in singular reflexives did not provide additional cues to antecedent identity above and beyond the contributions of number and syntactic position, the two nouns in each item set were chosen to have a similar gender bias, based in part on the norms in Kennison and Trofe (2003). 24 item sets contained a pair of male-biased nouns (as in Tables 1 and 2), and the remaining 24 contained a pair of female-biased nouns (e.g., *The legendary diva who sang with the graceful harpist*). In order to avoid biases from additional agreeing elements in the sentence, the verb inside the relative clause never overtly expressed agreement, and neither did the main clause verb used in the reflexive conditions. For all conditions, the subject noun phrase was followed by an adverbial that signaled the end of the relative clause.

In the agreement conditions the main verb was always a present tense agreeing form of *be* (*was* or *were*), followed by a predicative adjective and a four-word spillover region. For reflexives the main verb was always a non-agreeing, past tense verb that was followed immediately by a direct object reflexive. When the reflexive was singular it matched the gender bias of the two nouns in the sentence.

24 items contained stereotypically masculine head nouns, and 24 contained stereotypically female head nouns. As in the agreement conditions, the reflexive was followed by a four-word spillover region.

The 48 target items were mixed with 152 fillers, for a total of 200 sentences. In addition to the 24 ungrammatical target items, there were 12 unrelated ungrammatical fillers (containing an illicit NPI dependency), yielding an overall grammatical-to-ungrammatical ratio of 4.6:1. Half of the target items and half of the fillers were followed by a comprehension question. Across items, comprehension questions addressed various parts of the sentence; this was done in order to prevent participants from adopting superficial reading strategies that extracted the information needed to answer the comprehension questions without fully comprehending the sentence. The entire set of experimental stimuli can be found at <http://blogs.umass.edu/bwdillon>.

Offline judgments

One concern with investigating the effect of number mismatch on reflexive dependencies in English is that plural reflexives for singular antecedents are sometimes tolerated in situations where the speaker does not wish to commit to a particular gender for the antecedent. Many speakers report that sentences such as *the student hurt themselves during lunch break* may be acceptable in a colloquial register, but this effect appears to be subject to significant dialectal variation, and appears to be more widely accepted among younger speakers. However, for nouns that overtly signal the referent's gender, this option is much less commonly accepted, as in **the girl hurt themselves during lunch break*. In order to test whether the number mismatch in the present materials was reliably rejected by our target population, an offline judgment study was conducted with the 48 items from Experiment 1. These 48 target items were mixed with 100 fillers, and the materials were balanced so that across the experiment half of the sentences were ungrammatical. The anomalies in the fillers comprised a variety of different grammatical errors, including unlicensed NPIs and unlicensed verbal morphology (e.g. **will eating*). 12 naïve participants were asked to judge the acceptability of the sentences they read on a 7-point scale, where 7 was completely acceptable and 1 was completely unacceptable. Participants were instructed to judge the sentences with regard to their acceptability in colloquial speech. The results are presented in Table 3.

A three-way repeated measures ANOVA by participants revealed a significant main effect of grammaticality ($F(1, 11) = 34.0, p < .001$), as well as a significant interaction of grammaticality with distractor noun number ($F(1, 11) = 5.3, p < .05$). Resolving this interaction revealed no significant differences due to intrusion in pairwise comparisons. Thus there was no reliable evidence that the distractor's number had an effect on offline judgments for either dependency. Importantly, the main effect of grammaticality was highly significant for reflexives ($F(1, 11) = 31.7, p < .001$), and the magnitude of the grammaticality effect did not reliably differ across dependencies. These results confirm that in offline judgments, participants treat the reflexive

Table 3

Mean acceptability judgments and standard errors by participants in the Experiment 1 rating study. Values are on a 7-point scale where 7 is perfectly acceptable, and 1 is completely unacceptable.

	[+Gram, -Intr]	[+Gram, +Intr]	[-Gram, -Intr]	[-Gram, +Intr]
Agreement	5.36 (\pm .29)	5.08 (\pm .29)	3.00 (\pm .41)	3.32 (\pm .43)
Reflexives	5.60 (\pm .25)	5.68 (\pm .22)	3.08 (\pm .37)	3.24 (\pm .38)

and agreement anomalies in our stimuli as equally unacceptable.

Procedure

The 48 target item sets were distributed across 8 lists in a Latin Square design, and five participants were assigned to each list. Each list was randomized along with the filler sentences, subject to the constraint that no two experimental sentences were presented next to each other. The maximum number of characters allowed on a single line on the visual display was 142 characters, and all sentences in the experiment fit on one line. All sentences were presented in a 12-point fixed-width font (Courier), and all characters had a size of 9×16 pixels on the display. The resolution of the visual display was 1280×720 pixels on an LCD screen. Eye movements were recorded using an EyeLink 1000 tower-mount eye-tracker. Participants had binocular vision while movements were measured, but only the gaze of one eye was tracked. The tower was 32 in. from the visual display, giving participants approximately 5 characters per degree of visual angle. The eye-tracker sampled eye movements at 1000 Hz.

Prior to beginning the experiment, participants were familiarized with the apparatus and given four practice trials. While seated, participants' heads were immobilized using a chin rest and forehead restraint that was adjusted for comfort. Before the experiment, and whenever necessary throughout the experiment, the experimenter calibrated the eye-tracker with a 9-point display to ensure an accurate record of eye-movements across the screen. Participants initiated each trial by fixating on a marker at the beginning of the sentence. Once a fixation in the target region was recognized by the experimental software, the trial sentence was displayed all at once. Participants ended the presentation of the test sentence by indicating they had finished using a response pad. On trials with a question, the question was presented immediately after the test sentence, and participants indicated their response on the same response pad. Participants were allowed to take breaks at their discretion throughout the experiment. At a minimum, the experimenters asked the participants to take one short rest during the course of the experiment. After each break, participants were recalibrated to ensure accurate measurement of the eye movements.

Data analysis

Sentences in both the reflexive and the agreement conditions were divided into six regions of interest, as indicated in Tables 1 and 2. For all conditions the complex

subject noun phrase was divided into three regions: the head noun, together with its determiner and adjective (NP1), the relative clause complementizer and the embedded verb, and the distractor noun along with its determiner and adjective (NP2). The remainder of each sentence was divided into a pre-critical region, a critical region, and a spillover region. For agreement conditions the pre-critical region consisted of the main clause adverbial, the critical region consisted of the agreeing form of *be* and the following predicative adjective, and the spill-over region consisted of the remaining four words of the sentence. For reflexive conditions the pre-critical region consisted of the main clause adverbial and the main clause verb, the critical region consisted of the reflexive and the following preposition, and the spillover region was the remaining three words of the sentence. The extended analysis window for the critical region in the agreement conditions was adopted because of a high rate of skipping of the inflected auxiliary verb. An extended window was also adopted for the reflexive conditions in order to maximize the similarity of the critical reflexive region to the critical agreement region. We note, however, that similar patterns of results obtain when word-by-word regioning of the critical regions is used. Analysis for reflexive and agreement conditions proceeded separately in Experiment 1. The exception was the critical region, where the differences between agreement and reflexives were of theoretical interest.

We report three measures for each region of interest. The early measures reported here are the *first pass reading time* and the *probability of regression*. First pass reading time is calculated by summing all fixations in a region of interest after participants first enter the region until the first saccade out of that region (either to the right or the left). The probability of regression corresponds to the probability that a regression is initiated from a particular region before exiting that region to the right. We also report a late measure, *total time*, which is the total sum of all fixations in a particular region of interest, including first pass reading time and any time spent re-reading the region.

To analyze the probability of regression measure, we employed logistic mixed effects models (see Jaeger, 2008), as the dependent measure was categorical (i.e., presence or absence of a first-pass regression for a given region of interest). We used sum contrasts for experimental fixed effects (*grammaticality* and *intrusion*), and a fully-specified random effects structure, including random intercepts and slopes for all fixed effects by participants and by items (Baayen, Davidson, & Bates, 2008).

Results: Agreement

The comprehension question accuracy for the agreement conditions ranged between 88% and 93% accuracy, suggesting that participants successfully comprehended the experimental stimuli. There were no significant differences in accuracy between conditions.

The by-region reading times for first-pass and total time measures for the agreement conditions, as well as first-pass regression probabilities, are presented in Table 4.

Table 4

Table of means (in ms where applicable) for agreement conditions in Experiment 1 for first pass, total time, and probability of first-pass regression. Standard error by participants is shown in parentheses.

Agreement	NP1	NP2	Critical	Spillover
First pass				
<i>Gram, no int</i>	582 (30)	627 (37)	373 (18)	795 (42)
<i>Gram, int</i>	588 (32)	612 (34)	389 (17)	837 (48)
<i>Ungram, no int</i>	583 (36)	593 (31)	445 (23)	805 (51)
<i>Ungram, int</i>	594 (39)	639 (33)	439 (21)	847 (44)
Total time				
<i>Gram, no int</i>	867 (60)	955 (63)	579 (34)	1018 (46)
<i>Gram, int</i>	835 (51)	959 (50)	622 (40)	1059 (52)
<i>Ungram, no int</i>	891 (63)	971 (51)	811 (57)	1140 (70)
<i>Ungram, int</i>	881 (55)	1002 (70)	693 (35)	1074 (51)
Pr(Regression)				
<i>Gram, no int</i>	–	.18 (.03)	.13 (.03)	.65 (.04)
<i>Gram, int</i>	–	.20 (.03)	.17 (.02)	.67 (.04)
<i>Ungram, no int</i>	–	.18 (.03)	.25 (.04)	.71 (.04)
<i>Ungram, int</i>	–	.13 (.02)	.20 (.03)	.70 (.04)

Prior to the critical region, no significant effects for any of the experimental factors were observed in any measure.

A summary of the statistical analyses of the critical region is presented in Table 5. In the first-pass measures there was a main effect of grammaticality at the critical verb region ($F_1(1,39) = 16.5$, $p < .001$; $F_2(1,47) = 14.2$, $p < .001$). There were no significant effects in the spillover region. In the total time measure there was a significant effect of grammaticality ($F_1(1,39) = 25.5$, $p < .001$; $F_2(1,47) = 28.5$, $p < .001$), reflecting a slowdown for ungrammatical conditions relative to grammatical conditions. In addition there was an interaction between interference and grammaticality ($F_1(1,39) = 8.0$, $p < .01$; $F_2(1,47) = 7.9$, $p < .01$). Planned pairwise comparisons by participants revealed that this interaction was driven by a significant effect of interference for ungrammatical sentences ($t(39) = 2.7$, $p < .01$); the effect

of interference did not reach significance for comparisons between the grammatical conditions.

Additionally, analysis of the spill-over region for total times in the agreement conditions, revealed a significant main effect of grammaticality by participants ($F_1(1,39) = 4.6$, $p < .05$; $F_2(1,47) = 3.0$, $p < .1$).

In first-pass regression probabilities, logistic mixed effects models revealed a significantly greater probability of regression in the ungrammatical conditions relative to grammatical conditions at the critical agreeing verb ($\beta = .488$, Wald's $z = 2.2$, $p < .05$). Additionally, there was an interaction of grammaticality with interference ($\beta = -.878$, Wald's $z = -2.0$, $p < .05$). However, a model containing planned comparisons to resolve this interaction showed no significant differences due to interference within either grammatical or ungrammatical conditions. In the spillover region, there was a marginal effect of grammaticality on the probability of regression ($\beta = .263$, Wald's $z = 1.7$, $p < .1$).

Results: Reflexives

The comprehension question accuracy for the reflexive conditions ranged between 87% and 91%, suggesting that participants successfully comprehended the experimental stimuli. There were no significant differences observed between conditions.

The by-region reading times and regression probabilities for the reflexive conditions are presented in Table 6, and the results of the statistical analyses of the critical region are presented in Table 7. In first-pass measures no significant effects were observed prior to the critical region. At the critical reflexive region, there was a significant main effect of grammaticality, reflecting a slowdown in ungrammatical conditions ($F_1(1,39) = 19.5$, $p < .001$; $F_2(1,47) = 15.7$, $p < .001$). A main effect of grammaticality was also observed in total time measures at the critical reflexive region

Table 5

Summary of statistical analyses for agreement conditions in Experiment 1 at the critical agreeing verb region.

	By participants	By items	minF
First pass			
GRAM	$F_1(1,39) = 16.5$ MSE = 149159 $p < .001$	$F_2(1,47) = 14.2$ MSE = 212532 $p < .001$	$minF(1,86) = 7.62$ $p < .01$
INTR	$F_1(1,39) < 1$	$F_2(1,47) < 1$	$minF < 1$
GRAM \times INTR	$F_1(1,39) < 1$	$F_2(1,47) < 1$	$minF < 1$
Total time			
GRAM	$F_1(1,39) = 25.5$ MSE = 913241 $p < .0001$	$F_2(1,47) = 28.5$ MSE = 1362804 $p < .0001$	$minF(1,84) = 13.46$ $p < .001$
INTR	$F_1(1,39) = 2.4$ MSE = 57,198 $p < .2$	$F_2(1,47) = 2.1$ MSE = 105373 $p < .2$	$minF(1,86) = 1.12$ $p < .3$
GRAM \times INTR	$F_1(1,39) = 8.0$ MSE = 260996 $p < .01$	$F_2(1,47) = 7.9$ MSE = 413674 $p < .01$	$minF(1,85) = 3.96$ $p < .05$
	β	Wald's z	p
Pr(Regression)			
GRAM	$\beta = .488 (\pm .22)$	$z = 2.22$.027
INTR	$\beta = -.002 (\pm .20)$	$z = -.01$.990
GRAM \times INTR	$\beta = -.878 (\pm .43)$	$z = -2.0$.043

Table 6

Table of means (in ms where applicable) for reflexive conditions in Experiment 1 for first pass, total time, and probability of first-pass regression. Standard error by participants is shown in parentheses.

Reflexives	NP1	NP2	Critical	Spillover
First pass				
Gram, no int	583 (31)	619 (30)	299 (12)	650 (49)
Gram, int	579 (32)	645 (35)	295 (13)	659 (35)
Ungram, no int	570 (36)	621 (29)	351 (17)	628 (34)
Ungram, int	567 (32)	651 (29)	342 (16)	631 (27)
Total time				
Gram, no int	841 (49)	951 (55)	481 (20)	851 (48)
Gram, int	819 (47)	907 (48)	471 (24)	846 (42)
Ungram, no int	863 (56)	1023 (59)	588 (30)	918 (62)
Ungram, int	890 (59)	1040 (54)	580 (30)	882 (47)
Pr(Regression)				
Gram, no int	–	.19 (.03)	.14 (.03)	.73 (.04)
Gram, int	–	.18 (.03)	.16 (.03)	.77 (.04)
Ungram, no int	–	.16 (.03)	.09 (.02)	.77 (.03)
Ungram, int	–	.15 (.03)	.10 (.02)	.77 (.04)

($F_1(1,39) = 26.8, p < .0001$; $F_2(1,47) = 23.5, p < .0001$). In neither measure was there an interaction of grammaticality and interference. Although the interaction of interference and grammaticality did not reach significance, we nonetheless performed the same pairwise comparisons we used for the agreement conditions to maximize the chance of observing an intrusion effect. This analysis failed to show an effect of intrusion for either ungrammatical or grammatical sentences (all p 's $> .7$). Logistic mixed model analysis of the probability of regression measure revealed no significant effects.

There were no significant effects of the experimental manipulations on any of the measures at NP1. At the distractor NP position (NP2), however, there was a significant main effect of grammaticality on total reading times for reflexive conditions ($F_1(1,39) = 8.4, p < .01$; $F_2(1,47) = 6.8, p < .05$). There were no significant effects of interference or the interaction of interference and grammaticality at this region, nor were there effects in other measures.

Table 7

Summary of statistical analyses for reflexive conditions in Experiment 1 at the critical reflexive region.

	By participants	By items	minF'
First pass			
GRAM	$F_1(1,39) = 19.5$ MSE = 97,932 $p < .001$	$F_2(1,47) = 15.7$ MSE = 139902 $p < .001$	minF'(1,86) = 8.69
INTR	$F_1(1,39) < 1$	$F_2(1,47) < 1$	$p < .01$
GRAM \times INTR	$F_1(1,39) < 1$	$F_2(1,47) < 1$	minF' < 1
Total time			
GRAM	$F_1(1,39) = 26.8$ MSE = 468434 $p < .0001$	$F_2(1,47) = 23.5$ MSE = 730176 $p < .0001$	minF'(1,86) = 12.53
INTR	$F_1(1,39) < 1$	$F_1(1,47) < 1$	$p < .001$
GRAM \times INTR	$F_1(1,39) < 1$	$F_1(1,47) < 1$	minF' < 1
	β	Wald's z	p
Pr(Regression)			
GRAM	$\beta = -.148 (\pm .25)$	$z = -.59$.557
INTR	$\beta = -.273 (\pm .24)$	$z = -1.13$.261
GRAM \times INTR	$\beta = .093 (\pm .60)$	$z = .16$.877

Direct comparison of intrusion effect

Results showed a reliable effect of the number of an intruding noun for agreement, and no corresponding effect for reflexives. However, because the contrast between agreement and reflexives is of primary theoretical interest here, we directly tested the intrusion effect between agreement and reflexives by performing a $2 \times 2 \times 2$ repeated measures ANOVA that included the factors of dependency, grammaticality and interference. For total times this analysis revealed a three-way interaction that was significant by participants and marginal by items ($F_1(1,39) = 8.0, MSE = 134049, p < .01$; $F_2(1,47) = 4.0, MSE = 173282, p < .055$; $minF'(1,81) = 2.66, p = .11$). This interaction did not reach significance in either first-pass or probability of regression measures.

We resolved the critical three-way interaction in the total time measure by conducting planned pairwise comparisons on the size of the intrusion effect across the two grammatical dependencies. The intrusion effect was calculated by subtracting the reading time for non-intruding conditions from intruding conditions for total times in the critical region. This analysis revealed that there was a significantly larger intrusion effect for agreement in the ungrammatical conditions ($t(39) = -2.5, p < .02$; agreement $\mu = -118.6 \text{ ms} \pm 42.8$, reflexive $\mu = -7.8 \text{ ms} \pm 29.6$). There was no significant difference between agreement and reflexives in the grammatical conditions ($t(39) = 1.4, p < .3$; agreement $\mu = 43.0 \text{ ms} \pm 31.6$, reflexive $\mu = -10.1 \text{ ms} \pm 28.8$).

Discussion

For reflexive conditions, across all three measures reported, and for all other regions of interest (notably the distractor NP2), only grammaticality had a significant effect on eye movements. Ungrammatical conditions reliably led to longer reading times. Importantly, there were no significant effects of intrusion and no interaction of grammaticality with intrusion at any region or measure. As neither

inhibitory similarity-based interference nor facilitatory intrusion effects were observed for reflexive dependencies, there was no evidence for any impact of the structurally illicit noun phrase on the construction of the reflexive dependency.

The processing of reflexive conditions clearly contrasts with the processing of agreement conditions, which showed a qualitatively different pattern. Although ungrammaticality reliably slowed the processing of both agreement and reflexives, only agreement showed a reliable intrusion effect. This resulted in a significant interaction of grammaticality and interference at the critical agreement region for total times and probability of regression. Planned comparisons confirmed that this interaction was driven by intrusion in ungrammatical conditions. Intrusion led to shorter reading times (or fewer regressions) in ungrammatical sentences, replicating earlier findings for agreement in comprehension (Pearlmutter et al., 1999; Wagers et al., 2009).

Although the contrast between agreement and reflexives in total times is clear, first-pass measures did not so clearly indicate distinct processing profiles for the two dependencies. In first-pass reading times, agreement and reflexives both showed only a main effect of grammaticality. It is possible that this pattern provides evidence of an early processing stage where agreement and reflexives are both processed alike, and are not susceptible to intrusion. However, the fact that agreement does show an intrusion effect in another measure that taps early processing (probability of regression) does not support this interpretation. In addition, comprehenders may respond to early difficulty either by increased fixation duration or by launching a regressive saccade, and thus these two measures may trade off (Altmann, 1994; Rayner & Sereno, 1994). Indeed, a post hoc comparison of first-pass reading times in the critical region using a linear mixed effects model revealed a significant effect of regression ($t = -3.2$). Because of the presence of an early effect in the probability of regression measure, and the close relationship between regression times and reading times in these early measures, we do not believe that the first-pass reading time data support the conclusion that reflexives and agreement share a common, intrusion-free stage of early processing.

Experiment 2

Experiment 1 demonstrated that the computation of subject–verb agreement relations is reliably susceptible to intrusion. No corresponding intrusion effect was observed for reflexives. The main goal of Experiment 2 was to replicate this result for reflexives, which would reinforce the evidence for a lack of intrusion. In addition, we extended the design of Experiment 1 to include conditions where the head noun was plural rather than singular, to determine whether the same pattern is observed when the reflexive is bound by a plural NP. For agreement dependencies, it is known that the presence of intrusion interacts with the markedness of the features involved, such that no intrusion is observed when the agreeing subject is plural (Pearlmutter et al., 1999; Wagers et al., 2009). By testing

a wider range of feature combinations, we can address the possibility that intrusion might be observed for reflexives with a different combination of morphological features on the target and distractor nouns.

Participants

Thirty-two members of the University of Maryland community participated in Experiment 2 (23 females, mean age 21.3). Participants gave informed consent, and were either paid \$10 for their participation or received course credit. The experimental session, including set-up and calibration, lasted approximately 1 h.

Materials

The experimental materials and fillers were identical to those used in Experiment 1, with the exception that the four agreement conditions were removed. Instead, an additional factor of head noun number was introduced for the reflexive conditions, leading to a $2 \times 2 \times 2$ factorial design that crossed head noun number, interfering noun number, and verbal number. A full set of experiment conditions in Experiment 2 is found in Table 8.

Data analysis

Data analysis proceeded as in Experiment 1.

Results

Comprehension question accuracy averaged between 81% and 91% across the eight critical conditions, suggesting that participants performed the task satisfactorily and comprehended the critical sentences. Modeling revealed no significant differences in comprehension accuracy across the critical conditions.

Mean fixation durations and probability of regression at the regions used in the analysis are presented in Tables 9 and 10. In first pass and total time measures, modeling revealed a significant main effect of head noun number at NP1, reflecting a slow-down for plurals (first-pass; $F_1(1,31) = 12.9$, $p < .01$; $F_2(1,47) = 17.4$, $p < .001$); total time: $F_1(1,31) = 13.3$, $p < .01$; $F_2(1,47) = 12.5$, $p < .001$). Likewise, there was a significant main effect of interference at NP2, also reflecting a plural complexity effect (first pass: $F_1(1,31) = 4.5$, $p < .05$; $F_2(1,47) = 4.9$, $p < .05$; total time: $F_1(1,31) = 4.0$, $p < .06$; $F_2(1,47) = 4.9$, $p < .05$). Grammaticality effects were also observed in total time measures at the NP1 and NP2 regions (NP1: $F_1(1,31) = 22.5$, $p < .001$; $F_2(1,47) = 6.9$, $p < .05$; NP2: $F_1(1,31) = 15.6$, $p < .001$; $F_2(1,47) = 8.9$, $p < .01$).

A summary of the analyses at the critical reflexive region is presented in Table 11. Similar patterns were observed in both first pass and total time measures. We observed a main effect of grammaticality that was marginal in first pass times (first pass: $F_1(1,31) = 4.1$, $p < .06$; $F_2(1,47) = 5.9$, $p < .05$; total time: $F_1(1,31) = 13.7$, $p < .001$; $F_2(1,47) = 18.1$, $p < .001$). There was also an interaction of head number and grammaticality (first pass: $F_1(1,31) = 9.2$, $p < .01$; $F_2(1,47) = 8.8$, $p < .01$; total time: $F_1(1,31) = 18.1$, $p < .001$;

Table 8

Summary of reflexive conditions in Experiment 2. Regions included in the analysis are underlined.

Reflexive conditions for Experiment 2	
<i>Singular head, singular interferer, grammatical</i>	
1. <u>The new executive/</u> who oversaw/ <u>the middle manager/</u> apparently doubted/ <u>himself on/</u> <u>most major decisions</u>	
<i>Singular head, plural interferer, grammatical</i>	
2. <u>The new executive/</u> who oversaw/ <u>the middle managers/</u> apparently doubted/ <u>himself on/</u> <u>most major decisions</u>	
<i>Singular head, singular interferer, ungrammatical</i>	
3. <u>The new executive/</u> who oversaw/ <u>the middle manager/</u> apparently doubted/ <u>themselves on/</u> <u>most major decisions</u>	
<i>Singular head, plural interferer, ungrammatical</i>	
4. <u>The new executive/</u> who oversaw/ <u>the middle managers/</u> apparently doubted/ <u>themselves on/</u> <u>most major decisions</u>	
<i>Plural head, singular interferer, ungrammatical</i>	
5. <u>The new executives/</u> who oversaw/ <u>the middle manager/</u> apparently doubted/ <u>himself on/</u> <u>most major decisions</u>	
<i>Plural head, plural interferer, ungrammatical</i>	
6. <u>The new executives/</u> who oversaw/ <u>the middle managers/</u> apparently doubted/ <u>himself on/</u> <u>most major decisions</u>	
<i>Plural head, singular interferer, grammatical</i>	
7. <u>The new executives/</u> who oversaw/ <u>the middle manager/</u> apparently doubted/ <u>themselves on/</u> <u>most major decisions</u>	
<i>Plural head, plural interferer, grammatical</i>	
8. <u>The new executives/</u> who oversaw/ <u>the middle managers/</u> apparently doubted/ <u>themselves on/</u> <u>most major decisions</u>	

Table 9

Table of means (in ms where applicable) for reflexive conditions with a singular head noun in Experiment 2, for first pass, total time, and probability of first-pass regression. Standard error by participants is shown in parentheses.

Singular head	NP1	NP2	Critical	Spillover
First pass				
<i>Gram, sing</i>	586 (28)	611 (25)	292 (11)	676 (44)
<i>Gram, pl</i>	585 (29)	678 (32)	302 (13)	708 (42)
<i>Ungram, sing</i>	604 (30)	656 (31)	327 (12)	664 (44)
<i>Ungram, pl</i>	604 (28)	670 (32)	345 (16)	696 (38)
Total time				
<i>Gram, sing</i>	857 (63)	957 (50)	452 (24)	882 (61)
<i>Gram, pl</i>	872 (74)	1029 (56)	444 (20)	886 (56)
<i>Ungram, sing</i>	905 (57)	1050 (59)	574 (41)	952 (71)
<i>Ungram, pl</i>	956 (65)	1151 (64)	605 (32)	968 (54)
Pr(Regression)				
<i>Gram, sing</i>	–	.24 (.03)	.13 (.03)	.57 (.04)
<i>Gram, pl</i>	–	.19 (.03)	.09 (.02)	.47 (.05)
<i>Ungram, sing</i>	–	.19 (.02)	.15 (.03)	.64 (.05)
<i>Ungram, pl</i>	–	.21 (.03)	.12 (.02)	.67 (.04)

$F_2(1,47) = 13.8, p < .001$). Resolving this interaction with pairwise comparisons revealed that there was a significant effect of grammaticality in singular head noun conditions (first pass: $t(31) = 3.7, p < .001$; total time: $t(31) = 5.4, p < .001$), but no significant effects of grammaticality were observed when the head noun was plural (all p 's $> .6$). Logistic mixed effects modeling revealed no significant effects on the probability of regression at the critical region. Lastly, a main effect of grammaticality was observed in total reading times in the spillover region $F_1(1,31) = 6.1, p < .05$; $F_2(1,47) = 6.6, p < .05$).

Discussion

The most important finding from Experiment 2 is the replication of the processing profile for reflexives that was observed in Experiment 1. In no measure or region

Table 10

Table of means (in ms where applicable) reflexive conditions with a plural head noun in Experiment 2, for first pass, total time, and probability of first-pass regression. Standard error by participants is shown in parentheses.

Plural head	NP1	NP2	Critical	Spillover
First pass				
<i>Gram, sing</i>	681 (46)	651 (27)	311 (11)	742 (57)
<i>Gram, pl</i>	688 (38)	655 (35)	306 (11)	686 (42)
<i>Ungram, sing</i>	683 (39)	621 (24)	297 (10)	684 (52)
<i>Ungram, pl</i>	665 (35)	695 (25)	309 (15)	707 (46)
Total time				
<i>Gram, sing</i>	933 (62)	991 (46)	486 (21)	930 (86)
<i>Gram, pl</i>	967 (65)	1051 (66)	508 (26)	909 (79)
<i>Ungram, sing</i>	1046 (70)	1053 (56)	513 (26)	971 (70)
<i>Ungram, pl</i>	1077 (78)	1125 (57)	494 (27)	943 (77)
Pr(Regression)				
<i>Gram, sing</i>	–	.17 (.03)	.13 (.02)	.52 (.04)
<i>Gram, pl</i>	–	.22 (.03)	.15 (.02)	.58 (.05)
<i>Ungram, sing</i>	–	.24 (.03)	.15 (.03)	.59 (.05)
<i>Ungram, pl</i>	–	.18 (.04)	.18 (.04)	.61 (.05)

was there a reliable effect of distractor number, nor did it interact with any of the other factors. This strengthens support for the claim that reflexives are initially processed in a way that is unaffected by the feature content of a structurally illicit NP.

The results of Experiment 2 demonstrate the lack of intrusion effects for all combinations of head noun and distractor noun features. However, the results did reveal differences in the processing of ungrammatical reflexive dependencies across plural and singular head nouns. At the critical reflexive region comprehenders appeared to be less sensitive to the feature match between the reflexive and its antecedent when the head noun was plural. This apparent insensitivity to grammaticality is surprising in light of the fact that there was no comparable contrast in any other region that showed an effect of grammaticality for plural head noun conditions. These data could suggest that the process of recognizing feature mismatch in

Table 11
Summary of statistical analyses for Experiment 2 at the critical reflexive region.

	By participants	By items	minF
<i>First pass</i>			
HEAD	$F_1(1,31) = 1.6$ MSE = 7989 $p < .3$	$F_2(1,47) = 2.0$ MSE = 11,841 $p < .2$	$minF < 1$
PLUR	$F_1(1,31) = 1.4$ MSE = 5654 $p < .3$	$F_2(1,47) = 1.1$ MSE = 9279 $p < .4$	$minF < 1$
GRAM	$F_1(1,31) = 4.1$ MSE = 17,956 $p < .06$	$F_1(1,47) = 5.9$ MSE = 29,378 $p < .05$	$minF(1,68) = 2.41$ $p = .13$
HEAD × PLUR	$F_1(1,31) < 1$	$F_1(1,47) < 1$	$minF < 1$
HEAD × GRAM	$F_1(1,31) = 9.2$ MSE = 30,828 $p < .01$	$F_1(1,31) = 8.8$ MSE = 48,733 $p < .01$	$minF(1,75) = 4.51$ $p < .05$
GRAM × PLUR	$F_1(1,31) < 1$	$F_1(1,47) < 1$	$minF < 1$
HEAD × GRAM × PLUR	$F_1(1,31) < 1$	$F_1(1,47) < 1$	$minF < 1$
<i>Total time</i>			
HEAD	$F_1(1,31) < 1$	$F_2(1,47) = 1.3$ MSE = 31,161 $p < .3$	$minF < 1$
PLUR	$F_1(1,31) < 1$	$F_1(1,47) < 1$	$minF < 1$
GRAM	$F_1(1,31) = 13.7$ MSE = 349547 $p < .001$	$F_2(1,47) = 18.1$ MSE = 526399 $p < .0001$	$minF(1,69) = 7.79$ $p < .01$
HEAD × PLUR	$F_1(1,31) < 1$	$F_1(1,47) < 1$	$minF < 1$
HEAD × GRAM	$F_1(1,31) = 18.1$ MSE = 293592 $p < .001$	$F_2(1,47) = 13.8$ MSE = 473438 $p < .001$	$minF(1,78) = 7.82$ $p < .01$
GRAM × PLUR	$F_1(1,31) < 1$	$F_1(1,47) < 1$	$minF < 1$
HEAD × GRAM × PLUR	$F_1(1,31) = 1.9$ MSE = 26,115 $p < .2$	$F_2(1,47) = 2.2$ MSE = 47,782 $p < .2$	$minF(1,72) = 1.02$ $p < .4$
	β	Wald's z	p
<i>Pr(regression)</i>			
HEAD	$\beta = .314 (\pm .19)$	$z = 1.67$.096
PLUR	$\beta = -.070 (\pm .21)$	$z = -.34$.736
GRAM	$\beta = .185 (\pm .19)$	$z = .96$.338
HEAD × PLUR	$\beta = .067 (\pm .47)$	$z = .14$.885
HEAD × GRAM	$\beta = -.086 (\pm .46)$	$z = -.19$.851
GRAM × PLUR	$\beta = .032 (\pm .38)$	$z = .08$.933
HEAD × GRAM × PLUR	$\beta = .546 (\pm .87)$	$z = .64$.524

reflexive dependencies is sensitive to the markedness of the features involved, but this must remain speculative at this point. The empirical case for this conclusion is limited to a single region. In regions other than the critical region, such as the total times on NP1, a main effect of grammaticality was observed for singular and plural head nouns alike. A singular–plural contrast in comprehenders' ability to detect ungrammatical reflexive features is an interesting possibility that deserves further study, but at present this result must be treated with caution.

Computational model of the results

The results of Experiments 1 and 2 show that agreement and reflexives clearly differ with respect to their intrusion profiles. This distinction suggests that reflexives only use syntactic cues to retrieve their antecedents. This stands in contrast to verbal agreement, which appears to use agreement features as cues to retrieval of the subject.

However, it is logically possible that further contrasts between subject–verb agreement and reflexives, such as differences in the linear or structural position of the retrieval site, could be responsible for the observed difference in the intrusion profiles. Because there are explicit computational models of the retrieval process (Anderson, 1990; Lewis & Vasishth, 2005; Vasishth et al., 2008), it is informative to verify the formal predictions of the model under these different retrieval strategies and compare them to the experimental data. To derive these processing predictions, we used a variant of the ACT-R model described in Lewis and Vasishth (2005).

The ACT-R model is a cognitive architecture (Anderson & Lebiere, 1998) that has served as the basis for a successful computational model of syntactic parsing (Lewis & Vasishth, 2005; Vasishth et al., 2008). The core idea of the Lewis and Vasishth (2005) model is that syntactic parsing consists of passing linguistic constituents (*chunks*) between a passive working memory store and an active processing store in order to build syntactic structure.

Memory access works in a content-addressable fashion, using retrieval cues to access target memories when they are needed for processing. For further details of the ACT-R parsing model used here, see discussion in Dillon (2011) and supplemental materials provided at the first author's website (<http://blogs.umass.edu/bwdillon>).

By adopting an explicit model of memory retrieval, it is possible to compare quantitative model predictions to the results from Experiments 1 and 2. In order to facilitate this comparison, it is useful to clarify the linking assumptions we adopt to relate predicted retrieval latencies and online reading time measures. The most important assumption for present purposes is that any reading measures that are considered to index retrieval operations should be monotonically related to the retrieval latencies generated by the model. Although this is a useful starting assumption, it need not be the case: there are a number of processes that might obscure the relationship between retrieval time and observed reading time. Interpretive processes, reanalysis procedures, or error signals are likely important contributors to the observed behavioral measures. However, for current purposes, we assume with others (Lewis & Vasishth, 2005; Lewis et al., 2006; Vasishth et al., 2008) that whatever additional processes are reflected in the reading measures, they do not disrupt the monotonic relation between memory retrieval times and online reading time measures.

Modeling results

An important goal of the modeling experiments was to assess the impact of the unavoidable differences between agreement and reflexives, notably with respect to the parsing processes that precede the critical retrieval for the two dependencies. These processes could lead to differences in the baseline activation of the local subject for agreement and reflexives at the point of retrieval, as illustrated in Figs. 1 and 2. These figures show traces of the average ACT-R activation of the target and distractor NPs across the parse,

leading up to the critical retrieval. One notable difference between agreement and reflexives is evident: for reflexives, but not agreement, the grammatically appropriate target memory, i.e., the local subject, is reactivated immediately prior to the critical retrieval. This occurs due to the integration of the subject and the verb, and it gives the local subject a large activation advantage over the distractor at the subsequent retrieval of the reflexive's antecedent. This leads to a plausible alternative explanation of the experimental results: the lack of intrusion effects for reflexives may simply reflect differences in passive memory dynamics that are unrelated to the construction of the antecedent–reflexive dependency. If this activation difference confers a baseline bias in favor of the local noun immediately after processing the matrix verb, this may selectively eliminate intrusion effects for reflexive dependencies without the need to posit uniquely syntactic retrieval cues.

To determine whether the baseline activation differences of the subject noun were sufficient to account for the lack of intrusion effects with reflexives, the experimental conditions from Experiment 1 were modeled using an ACT-R parsing model. Rather than testing a single parameter setting, we considered the effects of a range of ACT-R parameter settings that span the values assumed in previous work. The advantage to this approach is that it allows the robustness of a prediction to be evaluated: strong model predictions hold across a range of parameter settings. The sole exception to this approach is the scaling parameter F , which was chosen to ensure that predictions of the model were on an appropriate time scale (in all simulations, $F = 2.0$). For each combination of parameters, 5000 Monte Carlo simulations were run. Each simulation included the full series of hypothesized retrievals, including random noise, and provided a prediction for the retrieval latency of the most probable memory chunk, as well as its identity (target or distractor).

The model provides two dependent measures of interest. The first is the proportion of retrievals that result in correct retrieval of the local subject. The second is a

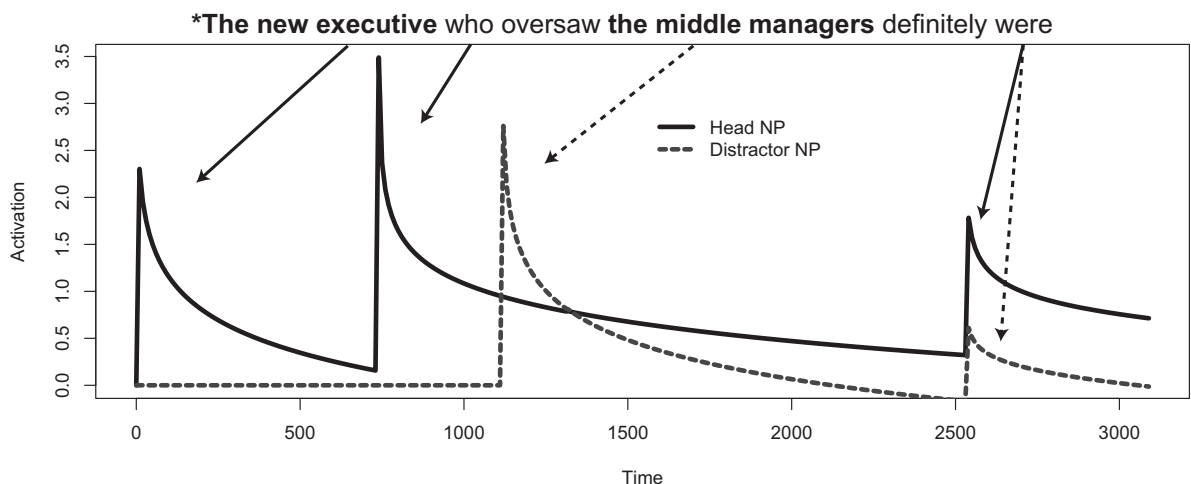


Fig. 1. Average activation for target (solid) and distractor (dashed) NPs for a sentence that shows intrusion at the agreeing verb. Incorrect retrievals of the distractor NP are reflected in the increased activation at the plural verb *were*.

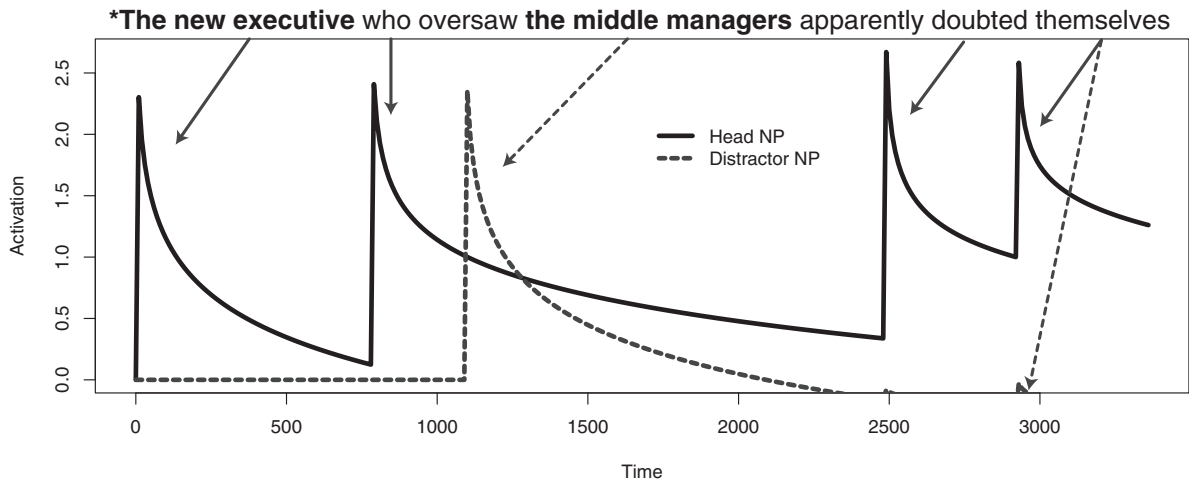


Fig. 2. Average activation for target (solid) and distractor (dashed) NPs for a sentence that shows intrusion at the reflexive. Incorrect retrievals of the distractor NP are reflected in the increased activation at the plural reflexive *themselves*.

Table 12

Mean percentage of retrieval error across parameter settings for agreement models, reflexive models with agreement cues (+Agr), and structured reflexive models (-Agr).

Model	[+Gram, -Intr]	[+Gram, +Intr]	[-Gram, -Intr]	[-Gram, +Intr]
Agreement	9.6%	5.1%	9.5%	21.9%
Reflexives (+Agr)	6.0%	3.9%	5.0%	11.4%
Reflexives (-Agr)	1.7%	1.7%	1.7%	1.7%

measure of the intrusion effect, which is derived from the predicated retrieval latencies that the model generates, regardless of which noun is retrieved. As in Experiment 1, the intrusion effect measures the predicted impact of the distractor's features on retrieval times. It is calculated within grammatical and ungrammatical pairs of conditions by subtracting the predicted retrieval time of the most probable noun for conditions with a plural distractor from conditions without a plural distractor.

We asked whether the reflexive data were best captured by agreement-based or uniquely syntactic cue sets for retrieval of the antecedent. In this experiment an agreement-based retrieval model for reflexive processing is one that used a mixture of syntactic and agreement features as cues to retrieve the target subject. Specifically, this included cues for *local subject* and *number*, and for singular reflexives, *gender*. For the syntactic retrieval model, we removed all morphological agreement cues to retrieval.

The empirical data of interest are the intrusion effects for reflexives observed in Experiments 1 and 2, as reflected in total reading times at the critical region. Total times were chosen because they provided the largest intrusion effect for agreement in Experiment 1, and so provided the best opportunity for an intrusion effect to be observed for reflexives, if there were any present. For each participant ($n = 40$ from Experiment 1, $n = 32$ from Experiment 2), the average intrusion effect for grammatical and ungrammatical pairs of reflexive conditions was calculated.

A comparison of the agreement-based reflexive model with a model of the subject-verb agreement retrieval process revealed differences in the rates of incorrect retrievals, as shown in Table 12. In addition, Fig. 3 shows the predicted size of the intrusion effect for each model. Despite the fact that agreement-based reflexive models used the same cues as the subject-verb agreement models to retrieve the local subject, the baseline activation differences between the two dependencies led to fewer incorrect retrievals and less intrusion for reflexives than for subject-verb agreement.

Despite the prediction for diminished intrusion for reflexives, however, a comparison between the predictions of agreement-based and syntactic retrieval models for reflexives shows that structured models provide a closer fit to our experimental data. Fig. 4 presents a comparison of the predicted intrusion effects and the observed intrusion effects in total reading times from Experiments 1 and 2. The lack of correlation between the retrieval cues and experimental manipulation in the structured access model led to no predicted differences in mean retrieval time across conditions. For this reason, the predicted intrusion effects were tightly centered around zero milliseconds. This was also seen in the error rates for structured access models in Table 12, which did not differ across conditions. This stood in contrast to agreement-based reflexive retrieval models, which predicted consistent intrusion effects for ungrammatical conditions.

The distribution of the empirical interference effect in Fig. 4 appears to better accord with the predictions of the syntactic retrieval model, suggesting that the data is better explained by the syntactic retrieval models. Interestingly, the modeling results also show that when agreement features are used in retrieval, intrusion effects are the most reliable behavioral prediction: there were no strong predictions for any sort of interference in the grammatical conditions. Because of the clear empirical contrast with agreement, and close fit with the predictions of the structured access model, we conclude that reflexives do not

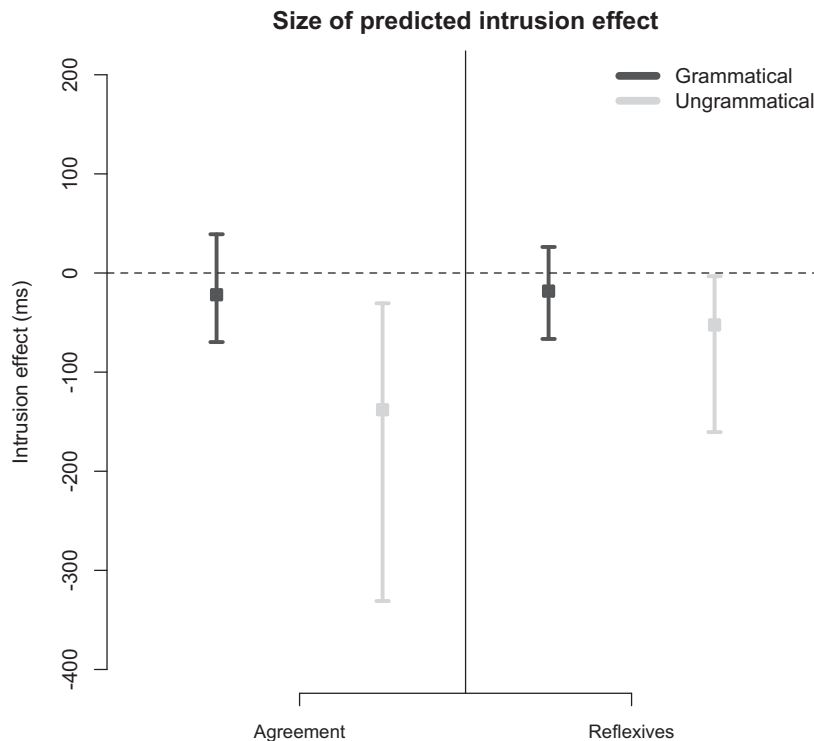


Fig. 3. Predicted intrusion effects for grammatical and ungrammatical subject–verb agreement and antecedent–reflexive dependencies using an identical set of subject retrieval cues, including agreement features. Error bars indicate 95% CI of the size of the intrusion effect across the range of parameter settings sampled.

recruit morphological features as retrieval cues when retrieving their antecedents. Instead, the retrieval of a reflexive's antecedent is only driven by syntactic cues to retrieval.

General discussion

Summary of results

In the present study we investigated the processing of agreement and reflexive dependencies across two eye-tracking experiments, and interpreted the results with the aid of a computational model of the processes underlying linguistic memory retrieval. The key question was whether the two dependencies showed the same intrusion profile, a prediction of parsing models that uniformly recruit agreement features as cues to memory retrieval in a content-addressable architecture. The eye-tracking studies showed a reliable difference in the intrusion profiles of agreement and reflexives during the critical subject retrieval. Fig. 5 presents a cross-experiment summary of the observed effects in total time measures at the critical region, demonstrating the contrast between agreement and reflexives. In Experiment 1, faster reading times were observed for ungrammatical subject–verb agreement dependencies when a distractor NP matched the features of the ungrammatical verb form, and no reliable effect of distractor NP number on grammatical agreement was observed. In contrast, in both Experiments 1 and 2 there was no impact of the distractor NP on the processing of reflexive

dependencies. Instead, across these two experiments reflexives were only sensitive to the feature match of the reflexive to the appropriate local subject. A direct comparison of the intrusion effects in Experiment 1 showed that there was a significant difference between the intrusion effect seen in agreement and the lack of an intrusion effect in reflexives. This contrast in intrusion profiles was observed even though the syntactic structure of the subject and the feature content of the distractor NP was identical for agreement and reflexive dependencies alike. These results are in line with previous findings on the sensitivity of these dependencies to the feature content of structurally illicit NPs, confirming the sensitivity of agreement to syntactically illicit NPs (Pearlmutter et al., 1999; Wagers et al., 2009) as well as confirming the resistance of reflexives to intrusion for number and gender features alike (Clackson et al., 2011; Nicol & Swinney, 1989; Sturt, 2003; Xiang et al., 2009).

In addition to demonstrating a consistent contrast in intrusion profiles for agreement and reflexives, a number of further generalizations are suggested by our data. In Experiment 1 the intrusion effect for agreement was limited to configurations with a singular head and a plural distractor, mirroring contrasts observed in production and comprehension findings (Bock & Miller, 1991; Eberhard et al., 2005; Wagers et al., 2009). In Experiment 2 we replicated the lack of intrusion effects for reflexives, and showed that this generalization held even when the subject head noun was plural. However, although Experiment 2 showed no evidence for intrusion for reflexives, the

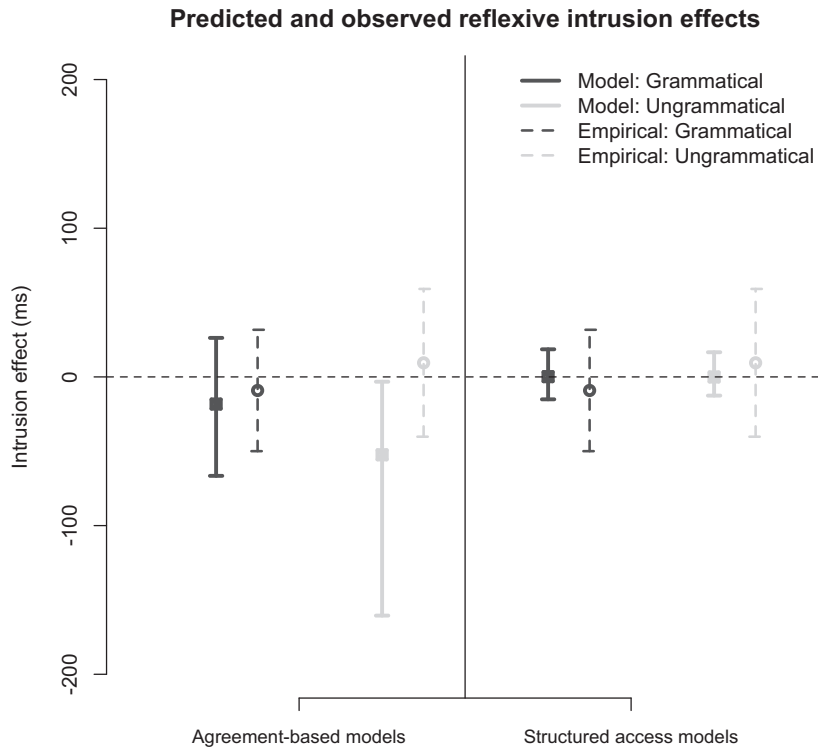


Fig. 4. Comparison of predicted intrusion effects (solid) of the retrieval model for reflexives and observed reflexive intrusion effects (dashed) from Experiments 1 and 2 (by participants, $n = 72$). For model predictions, error bars indicate 95% CI of the size of the intrusion effect across the range of parameter settings sampled; for experimental data, error bars reflect 95% CI for the intrusion effect by participants.

Summary of intrusion effects in Experiments 1 and 2

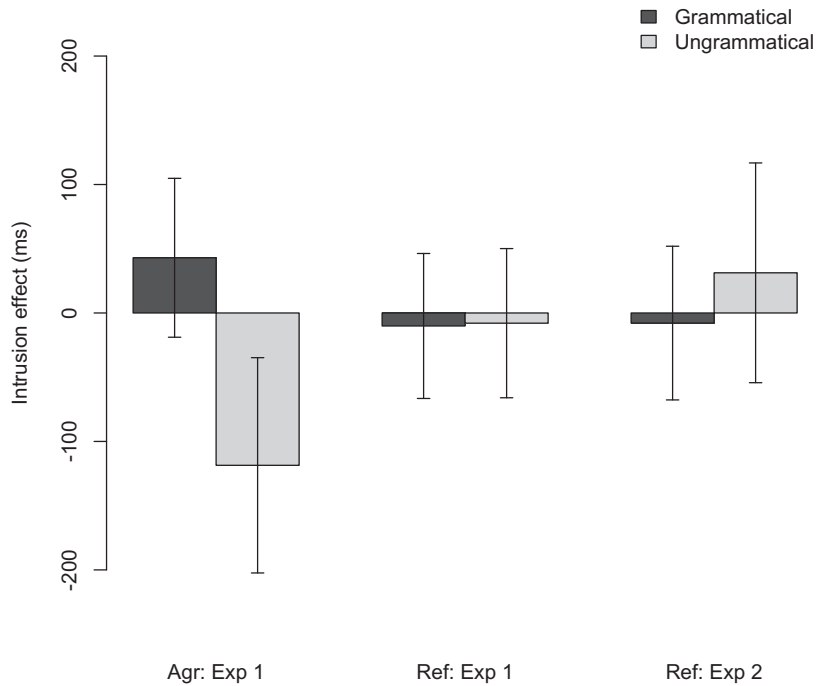


Fig. 5. Intrusion effects (in ms) observed for in total time measures for agreement and reflexive dependencies, across experiments. Error bars reflect 95% CI by participants.

results surprisingly revealed no effect of ungrammaticality for reflexives with a plural head noun at the critical reflexive region. Instead, the grammaticality effect was observed in increased reading times for ungrammatical conditions at spillover regions and in total reading times on the subject noun. This result may suggest that participants are slower to detect a feature-mismatched reflexive when it is bound by a plural subject than when it is bound by a singular subject, such that the effect of grammaticality only appears in the spillover region. If true, this is consistent with our hypothesis that reflexives do not use morphological cues in retrieval. If the morphological cues were deployed during retrieval of the subject, then the feature mismatch should have been detected at the reflexive. Although this is a potentially interesting finding, the experiments presented here were not designed to test this possibility, and so any conclusions drawn from this pattern are tentative at best. We leave further investigation of this effect to future research.

We also conducted computational simulations of the subject retrieval processes that are engaged during the construction of agreement and reflexive dependencies. These simulations revealed two main findings. First, the model predicted reflexives to show a diminished intrusion effect relative to agreement simply because of the post-verbal position of the reflexive. Specifically, the correct main clause subject is independently reactivated immediately prior to the critical reflexive, due to thematic binding of the subject and verb. This gives the main clause subject an activation advantage that reduces the impact of intrusion from distractor NPs during retrieval of the reflexive's antecedent. However, the main result of the simulations is that although reflexives are predicted to show less intrusion than agreement, the empirical results from Experiments 1 and 2 are more likely on a structured access model of the retrieval of the reflexive's antecedent. Access mechanisms that use agreement cues in retrieval do still predict intrusion effects for reflexives in this environment, which were not observed in any of our experiments.

Structure in linguistic memory access

The present results suggest that there is no uniform mapping between overt linguistic features and the retrieval cues used to parse a given linguistic dependency. Instead, it appears that selection of the retrieval cues is determined in part by the underlying grammatical function of the dependency being processed. Our present results suggest that agreement features are deployed as retrieval cues during the comprehension of subject-verb agreement, leading to the characteristic and widely-observed agreement intrusion effects (Clifton et al., 1999; Pearlmutter et al., 1999; Wagers et al., 2009). Despite similar surface correlations of the overt agreement features between a reflexive and its antecedent, our modeling and experimental results suggest that retrievals associated with the processing of reflexives rely primarily on syntactic cues. It has been argued that the binding dependency between a reflexive and its antecedent is established without regard to agreement features (Baker, 2008; Kratzer, 2009). If this is correct, then the present results may reflect

the grammatical primacy of the referential relation over morphological features for binding dependencies.

The core observation that agreement and reflexive dependencies treat morphological features in qualitatively different manners can be implemented in a number of different ways. In the present simulations, we modeled this as a narrowly syntactic set of retrieval cues in the case of reflexive dependencies. There are a number of alternative possibilities, however. For example, Van Dyke and McElree (2011) discuss the importance of the relative weights of different retrieval cues (cue combinatorics). They suggest that syntactic and semantic cues may be combined in a non-linear fashion such that syntactic cues effectively gate access to other cues in the set of retrieval cues. Our data may also be modeled in this fashion: syntactic cues may gate access to morphological cues for reflexive dependencies, whereas the two sets of cues are combined in a linear fashion for subject-verb agreement dependencies. Either implementation retains the core insight of the present results, namely that the agreement features are not active in retrieval of the reflexive's antecedent in the same way that they are for subject-verb agreement.

The present results are also consistent with models that adopt qualitatively different retrieval mechanisms for different linguistic dependencies (Dillon, 2011; Phillips et al., 2011). If reflexive dependencies are parsed using a (serial) search mechanism that operates over a complete syntactic tree (Berwick & Weinberg, 1984; Dillon, 2011), then it is not surprising that the agreement features of structurally illicit noun phrases do not interfere in the process of finding an antecedent. Such a mechanism may be distinguished from direct access mechanisms on the basis of time course data: if the time course of memory retrieval is impacted by distance for reflexive dependencies, then the search mechanism implementation would be supported. However, the present results do not decide between these different implementations of retrieval operation engaged by argument reflexives in our data.

It may also be the case that the source of the syntactic retrieval process observed for reflexives is linked to the argument structure of the verb, which some theoretical accounts of binding relationships hold to be of central importance in describing the distribution of reflexive anaphors (e.g., Pollard & Sag, 1992; Steedman, 2000). Although our computational simulations suggested that the post-verbal position of the reflexive was not sufficient to capture the empirical difference between agreement and reflexives, it is possible that the post-verbal position of the reflexive is important because it allows the use of argument structure information in retrieval or because the subject is maintained in the focus of attention while the predicate is processed (as argued by King, Andrews, & Wagers, 2012). These alternative accounts predict that the grammatically accurate retrieval of a reflexive's antecedent should be limited to coargument reflexives that are adjacent to the verb. On these accounts, we should observe intrusion effects with non-argument reflexives, preverbal reflexives, or long-distance reflexives. On the other hand, if comprehenders have a reliable mechanism for retrieving grammatically licit subject NPs while constructing a binding

dependency online, then we do not predict the presence of intrusion effects for these types of anaphors.

One possible concern is that the failure to observe intrusion effects for reflexive dependencies reflects an interaction between the activation advantage of the local subject subsequent to the verb and the diminished syntactic prominence of the distractor NP (which was in an embedded object position). We believe this to be an unlikely possibility for a number of reasons. One important fact that casts doubt on this explanation of the current data is that the lack of intrusion effects we observe is consistent with a broad literature on the processing of reflexives that consistently fails to find intrusion effects (Clackson et al., 2011; Clifton et al., 1999; Nicol & Swinney, 1989; Sturt, 2003; Xiang et al., 2009). Importantly, a number of these studies adopt designs that place distractor NPs in syntactically prominent positions: Sturt (2003; Experiment 1) and Nicol and Swinney (1989) used sentence-initial, topical subjects as distractors, and Xiang et al. (2009) used embedded subjects as the distractors. These studies likewise failed to find an intrusion effect for reflexives, suggesting that the present results cannot be uniquely attributed to the combination of decreased saliency and linear order. In addition, this explanation relies on there being an interaction between distractor saliency and linear order that does not necessarily follow from the model of the retrieval mechanism we adopt here. Neither linear order nor saliency alone can account for the lack of intrusion in reflexives: in our experimental design, the syntactic position of the distractor was matched between agreement and reflexives, and the computational simulations show that a simple additive relation between linear order and saliency effects on activation is not enough to eliminate the intrusion effect. Without a specific reason to assume interactivity, it seems that an additive relationship between these two variables is a plausible default hypothesis.

It might be argued that Badecker and Straub's (2002) study presents a potential counterexample to the predictions of structured access for reflexives. Badecker and Straub tested whether interference from illicit antecedents is observed in an online measure such as self-paced reading. They carried out four reading studies that involved reflexives or reciprocals; an example of one comparison is given in (3).

-
- (3) a. Jane thought that Bill owed himself a second chance.
 b. John thought that Bill owed himself a second chance.
-

In two of their four self-paced reading experiments with reflexives, Badecker and Straub found evidence for what they called the *multiple-match effect*, a slowdown in reading times following the reflexive in (3b) relative to (3a). For reflexives, this was either observed several regions past the critical reflexive (Experiment 3), or in pooled reading times at the four regions following the critical anaphor (Experiment 4). These results are similar to those reported by Chen et al. (2011) and Patil et al. (2011), who

also reported processing difficulty for processing reflexive anaphors in grammatical sentences with multiple, feature-matched NPs. Although these results do not provide evidence for the facilitatory intrusion effects predicted by feature-based retrieval mechanisms, the processing slowdowns are consistent with similarity-based interference effects that arise in these models when gender cues are used at retrieval (Chen et al., 2011; Patil et al., 2011). However, we do not believe these results provide strong evidence for the use of non-structural cues in retrieval of a reflexive's antecedent, because inhibitory interference effects of this sort need not be due to retrieval interference. For example, a feature-overwriting process that degrades the memory fidelity of individual memory chunks when their representations overlap in feature content would predict similar patterns of difficulty, without requiring the use of non-structural cues to retrieval (Gordon et al., 2001, 2002, 2004; Nairne, 1988, 1990). If the multiple-match effect reflects such a feature-overwriting process, then these findings do not support strong conclusions about the cues used to retrieve the antecedent of a reflexive. Unlike inhibitory interference effects, facilitatory intrusion effects receive a relatively unambiguous interpretation in the context of cue-based parsing models; they reflect the use of non-structural cues at retrieval that lead to temporary illusions of grammaticality. Thus we believe that the strongest evidence that comprehenders use non-structural cues to retrieval, and thereby genuinely retrieve illicit antecedents during online comprehension, is the presence of a facilitatory intrusion effect. In addition, we note that the prediction of intrusion effects was the clearest prediction of agreement-based retrieval models for reflexives in our simulations (for further discussion see Dillon, 2011). We are unaware of any experimental data that suggest that adult comprehenders are specifically susceptible to intrusion effects while processing reflexives, supporting the conclusion that parsing English reflexives involves a retrieval process that uses only syntactic cues to retrieve the grammatically appropriate local subject.

Differences between agreement and reflexives

We suggested that the differences in the use of agreement features as retrieval cues for agreement and reflexive dependencies reflect representational differences between the two dependencies. Indeed, the results of Experiments 1 and 2, are consistent with grammatical models that suggest that agreement in binding dependencies is established only subsequent to the construction of a referential dependency (Baker, 2008; Kratzer, 2009). Another possibility is that reflexive and agreement dependencies are computed at distinct levels of representation in the course of online processing. For example, reflexives may be computed at interpreted levels of linguistic representation (perhaps as constraints on interpretation; Chomsky & Lasnik, 1993; Jackendoff, 1972; Wasow, 1972), and that memory within these levels of interpretation employs distinct sets of retrieval cues. This idea is similar to the account advanced by Eberhard et al. (2005), who suggest that the influence of interpreted levels of representation helps explain the

greater influence of notional number on the expression of pronominal agreement more generally.

Alternatively, it is possible that the differences between reflexive and agreement dependencies may reflect differences in the predictability of the dependency. Wagers et al. (2009) suggested that the retrieval operation for agreement dependencies is a sort of error-driven processing that is engaged only when top-down expectations about the agreement features of the verb conflict with the bottom-up input. This account offers one explanation of why intrusion occurs only in ungrammatical sentences, where the verb form does not have the expected feature content. Reflexives cannot be anticipated as reliably as agreement, and so it is less likely that comprehenders should have strong top-down expectations about the feature content of an anaphor in direct object position before encountering the reflexive in the input. In the absence of any top-down expectation, memory retrieval is the only way of accessing the antecedent of the reflexive, and so comprehenders may adopt a more conservative strategy for accessing the appropriate local subject.

Conclusion

The results of the present study suggest that despite superficial similarity, agreement and reflexive dependencies show qualitatively different intrusion profiles. While agreement is reliably susceptible to intrusion from structurally illicit distractors, reflexive dependencies resist intrusion in identical environments. This pattern of results suggests that morphological agreement constraints are not uniformly implemented as highly ranked agreement retrieval cues. Instead, although this appears to be the case for subject–verb agreement, the lack of intrusion for reflexives suggests that retrieval of the reflexive's antecedent is guided solely by syntactic cues. This provides further evidence that comprehenders can deploy a structured access mechanism to recover information from linguistic memory for some linguistic dependencies. These findings are informative with respect to the question of how grammatical constraints relate to retrieval cues, and suggest that the implementation of feature concord constraints in online processing depends on the grammatical function of the linguistic dependency that they express.

Acknowledgments

This work was supported in part by NSF IGERT DGE-0801465 to the University of Maryland and by NSF BCS-0848554 to CP. We are particularly grateful to Rick Lewis for his help in performing the computational simulations reported here. We are grateful to Chuck Clifton, Lyn Frazier, Norbert Hornstein, Jeff Lidz, Umesh Patil, Michael Shvartsman, Adrian Staub, Shravan Vasishth and Matt Wagers for useful discussions of the material presented here.

References

Altmann, G. T. M. (1994). Regression-contingent analyses of eye movements during sentence processing: Reply to Rayner and Sereno. *Memory & Cognition*, 22, 286–290.

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- 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.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59, 390–412.
- Baker, M. (2008). *The syntax of agreement and concord*. Cambridge, UK: Cambridge University Press.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In *The crosslinguistic study of sentence processing* (pp. 3–73).
- Berwick, R., & Weinberg, A. (1984). *The grammatical basis of linguistic performance*. Cambridge: MIT Press.
- Bock, K., & Miller, C. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–93.
- Chen, Z., Jaeger, L., & Vasishth, S. (2011). How structure-sensitive is the parser? Evidence from Mandarin Chinese. In B. Stolterfoht & S. Featherston (Eds.), *Current work in linguistic evidence: The fourth Tübingen meeting. Studies in generative grammar*. Berlin: Mouton de Gruyter.
- Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. In J. Jacobs, A. Von Stechow, & W. Sternefeld (Eds.), *Syntax: An international handbook of contemporary research* (Vol. 1). Berlin: Mouton de Gruyter.
- Chomsky, N., & Miller, G. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. II). New York: John Wiley & Sons.
- Clackson, K., Felsler, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, 65, 128–144.
- Clifton, C., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Rivista di Linguistica*, 11, 11–39.
- Corbett, G. (2006). *Agreement*. Cambridge: Cambridge University Press.
- Den Dikken, M. (2001). "Plurilinguals", pronouns, and quirky agreement. *The Linguistic Review*, 18, 19–41.
- Dillon, B. (2011). *Structured access in comprehension*. Doctoral dissertation. University of Maryland, College Park.
- Drenhaus, H., Frisch, S., & Saddy, D. (2005). Processing negative polarity items: When negation comes through the backdoor. In S. Kepser & M. Reis (Eds.), *Linguistic evidence-empirical, theoretical, and computational perspectives*. Berlin: Mouton de Gruyter.
- Eberhard, K., Cutting, J., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, 112, 531–559.
- Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Availability versus accessibility. *Journal of Memory and Language*, 56, 357–383.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during sentence processing. *Psychological Science*, 13, 425–430.
- Gordon, P. C., Hendrick, R., & Levine, W. (2002). Memory-load interference in syntactic processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1411–1423.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, 51, 97–114.
- Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 32, 1304–1321.
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Jaeger, F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Kayne, R. (1989). Notes on English agreement. *Central Institute of English and Foreign Languages Bulletin*, 1, 46–67.
- Kennison, S., & Trofe, J. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32, 355–378.
- Kimball, J., & Aissen, J. (1971). I think, you think, he think. *Linguistic Inquiry*, 2, 241–246.
- King, J., Andrews, C., & Wagers, M. (2012). Do reflexives always find a good antecedent for themselves? Poster at the 25th annual meeting of the CUNY conference on human sentence processing, New York, NY, March 14–16.

- Kratzer, A. (2009). Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry*, 40, 187–237.
- Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375–419.
- Lewis, R., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, 10, 447–454.
- Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb–phrase ellipsis. *Journal of Memory and Language*, 58, 879–906.
- Martin, A. E., & McElree, B. (2009). Memory operations that support sentence comprehension: Evidence from verb–phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 35, 1231–1239.
- 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. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, 16, 343–352.
- Nairne, J. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251–269.
- Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5–19.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25, 273–285.
- Pearlmutter, N., Garnsey, S., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, 41, 427–456.
- Patil, U., Vasishth, S., & Lewis, R. (2011). Early retrieval interference in syntax-guided antecedent search. Oral presentation at the 24th annual meeting of the CUNY conference on human sentence processing, Stanford, CA, March 24–26.
- Phillips, C., Wagers, M., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the interference, syntax and semantics* (Vol. 37). Bingley, UK: Emerald Publications.
- Pollard, C., & Sag, I. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry*, 23, 261–303.
- Rayner, K., & Sereno, S. C. (1994). Regressive eye movements and sentence parsing: On the use of regression-contingent analyses. *Memory & Cognition*, 22, 281–285.
- Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, 49, 1–46.
- Steedman, M. (2000). *The syntactic process*. Cambridge: MIT Press.
- 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. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 407–430.
- Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49, 285–316.
- Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence processing. *Journal of Memory and Language*, 55, 157–166.
- Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, 65, 247–263.
- Vasishth, S., Brüssow, S., Lewis, R., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, 32, 685–712.
- Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, 68, 13–29.
- Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject–verb agreement in Spanish and English: Differences in the role of conceptual factors. *Cognition*, 61, 261–298.
- Wagers, M. (2008). *The structure of memory meets memory for structure in linguistic cognition*. Doctoral dissertation. University of Maryland, College Park.
- Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237.
- Wasow, T. (1972). *Anaphoric relations in English*. Doctoral dissertation. Massachusetts Institute of Technology.
- Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing across dependency types: ERP evidence. *Brain and Language*, 108, 40–55.